

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:01:44 ; Search time 57.44 Seconds
(without alignments)
370.748 Million cell updates/sec

Title: US-09-495-823-7

Perfect score: 3012

Sequence: 1 MAPRCAGHPPPSPQACVC.....VPRYPKPRSNRLNGV 550

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1426.5	47.4	533	1 ARSB_HUMAN	P15848 homo sapien
2	1414.5	47.0	535	1 ARSB_FELCA	P33727 felis silve
3	1333	44.3	473	1 ARSB_RAT	P50430 rattus norv
4	864	28.7	285	1 ARSB_MOUSE	P50429 mus musculu
5	430.5	14.3	591	1 ARSF_HUMAN	P54793 homo sapien
6	429	14.2	535	1 ARS_PSEAE	P51691 pseudomonas
7	422.5	14.0	522	1 GA6S_HUMAN	P34059 homo sapien
8	416.5	13.8	567	1 ARS_STRPU	P50473 strongyloce
9	413	13.7	583	1 STS_HUMAN	P08842 homo sapien
10	407.5	13.5	589	1 ARSE_HUMAN	P51690 homo sapien
11	403	13.4	577	1 STS_RAT	P51589 rattus norv
12	399	13.2	593	1 ARSD_HUMAN	P51689 homo sapien
13	394	13.1	551	1 ASLA_ECOLI	P25549 escherichia
14	392.5	13.0	624	1 STS_MOUSE	P50427 mus musculu
15	378	12.5	507	1 ARSA_HUMAN	P51289 homo sapien
16	364	12.1	506	1 ARS_KLEAE	P50428 mus musculu
17	356	11.8	464	1 ARS_HEMPU	P20713 klebsiella
18	349.5	11.6	551	1 ARS_HEMPU	P14000 hemocitrot
19	324	10.8	560	1 YDEN_ECOLI	P77318 escherichia
20	250	8.3	559	1 GL6S_CAPHI	P50426 capra hircu
21	248.5	8.3	552	1 GL6S_HUMAN	P51586 homo sapien
22	247	8.2	649	1 ARS_VOLCA	Q10723 volvox cart
23	241	8.0	497	1 YIDJ_ECOLI	P31447 escherichia
24	231	7.7	647	1 ARS_CHLRE	P14217 chlamydomon
25	214	7.1	502	1 SPHM_HUMAN	P51688 homo sapien
26	205.5	6.8	550	1 IDS_HUMAN	P22304 homo sapien
27	197	6.5	563	1 IDS_MOUSE	Q08890 mus musculu
28	184.5	6.1	512	1 BETC_RHIME	O69787 rhizobium m
29	135	4.5	647	1 YC46_HAEIN	P44135 haemophilus
30	122.5	4.1	638	1 YQGS_BACSU	P54496 bacillus su
31	116	3.9	553	1 HISS_EMENI	Q9P4P9 emericella
32	106.5	3.5	437	1 AM3E_ORYSA	P27934 oryza sativ
33	106.5	3.5	763	1 MDOB_ECOLI	P39401 escherichia

RESULT 1	ARSB_HUMAN	105	3.5	922	1	NRPI_RAT	Q9qwj9
ID	ARSB_HUMAN	AC	P15848;	STANDARD;	PRT;	533 AA.	rattus norv
DT	01-APR-1990 (Rel. 14, Created)	DT	01-APR-1990 (Rel. 14, Last sequence update)				rattus norv
DT	30-MAY-2000 (Rel. 39, Last annotation update)	DT	30-MAY-2000 (Rel. 39, Last annotation update)				rhizobium m
DE	Arylsulfatase B precursor (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-4-sulfatase) (G4S).	DE	Arylsulfatase B precursor (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-4-sulfatase) (G4S).				pedobacter
GN	ARSB.	GN	ARSB.				mus musculu
OS	Homo sapiens (Human).	OS	Homo sapiens (Human).				murine coro
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				cucurbita m
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				haemophilus
OX	NCBI_TaxID=9606;	OX	NCBI_TaxID=9606;				homo sapien
RN	[1]	RN	[1]				salmonella
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				homo sapien
RX	MEDLINE=90153994; PubMed=2303452;	RX	MEDLINE=90153994; PubMed=2303452;				
RA	Peters C., Schmidt B., Rommerskirch.W., Rupp K., Zuehlendorf M.,	RA	Peters C., Schmidt B., Rommerskirch.W., Rupp K., Zuehlendorf M.,				
RA	Vingron M., Meyer H.E., Pohlmann R., von Figura K.;	RA	Vingron M., Meyer H.E., Pohlmann R., von Figura K.;				
RT	"Phylogenetic conservation of arylsulfatases. cDNA cloning and	RT	"Phylogenetic conservation of arylsulfatases. cDNA cloning and				
RT	expression of human arylsulfatase B.";	RT	expression of human arylsulfatase B.";				
RL	J. Biol. Chem. 265:3374-3381(1990).	RL	J. Biol. Chem. 265:3374-3381(1990).				
RN	[2]	RN	[2]				
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90152677; PubMed=1968043;	RX	MEDLINE=90152677; PubMed=1968043;				
RA	Schuchman E.H., Jackson C.E., Desnick R.J.;	RA	Schuchman E.H., Jackson C.E., Desnick R.J.;				
RA	"Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-	RA	"Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-				
RT	length cDNA, and regions of amino acid identity with arylsulfatases A	RT	length cDNA, and regions of amino acid identity with arylsulfatases A				
RT	and C.";	RT	and C.";				
RL	Genomics 6:149-158(1990).	RL	Genomics 6:149-158(1990).				
RN	[3]	RN	[3]				
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93332648; PubMed=7687847;	RX	MEDLINE=93332648; PubMed=7687847;				
RA	Modarelli S., Rupp K., von Figura K., Peters C.;	RA	Modarelli S., Rupp K., von Figura K., Peters C.;				
RT	"Structure of the human arylsulfatase B gene.";	RT	"Structure of the human arylsulfatase B gene.";				
RL	Biol. Chem. Hoppe-Seyler 374:327-335(1993).	RL	Biol. Chem. Hoppe-Seyler 374:327-335(1993).				
RN	[4]	RN	[4]				
RP	SEQUENCE OF 1-104 FROM N.A.	RP	SEQUENCE OF 1-104 FROM N.A.				
RX	MEDLINE=92028992; PubMed=1930244;	RX	MEDLINE=92028992; PubMed=1930244;				
RA	Litjens T., Morris C.P., Gibson G.J., Beckmann K.R., Hopwood J.J.;	RA	Litjens T., Morris C.P., Gibson G.J., Beckmann K.R., Hopwood J.J.;				
RT	"Human N-acetylgalactosamine-4-sulphatase: protein maturation and	RT	"Human N-acetylgalactosamine-4-sulphatase: protein maturation and				
RT	isolation of genomic clones.";	RT	isolation of genomic clones.";				
RL	Biochem. Int. 24:209-215(1991).	RL	Biochem. Int. 24:209-215(1991).				
RN	[5]	RN	[5]				
RP	2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.	RP	2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.				
RX	MEDLINE=95354208; PubMed=7628016;	RX	MEDLINE=95354208; PubMed=7628016;				
RA	Schmidt B., Selmer T., Ingendoh A., von Figura K.;	RA	Schmidt B., Selmer T., Ingendoh A., von Figura K.;				
RT	"A novel amino acid modification in sulfatases that is defective in	RT	"A novel amino acid modification in sulfatases that is defective in				
RT	multiple sulfatase deficiency.";	RT	multiple sulfatase deficiency.";				
RL	Cell 82:271-278(1995).	RL	Cell 82:271-278(1995).				
RN	[6]	RN	[6]				
RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).	RP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).				
RX	MEDLINE=97184692; PubMed=9032078;	RX	MEDLINE=97184692; PubMed=9032078;				
RA	Bond C.S., Clements P.R., Ashby S.J., Collyer C.A., Harrop S.J.,	RA	Bond C.S., Clements P.R., Ashby S.J., Collyer C.A., Harrop S.J.,				
RA	Hopwood J.J., Guss J.M.;	RA	Hopwood J.J., Guss J.M.;				
RT	"Structure of a human lysosomal sulfatase.";	RT	"Structure of a human lysosomal sulfatase.";				

ALIGNMENTS


```

ARBSE_RAT      STANDARD;          PRT;      473 AA.
ID  ARSB_RAT
AC  P50430;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  15-JUL-1998 (Rel. 36, Last annotation update)
DE  Arylsulfatase B (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-
DE  4-sulfatase) (G4S) (Fragment).
GN  ARSB.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
SEQUENCE FROM N.A.
RC  STRAIN=SPAGUE-DAWLEY; TISSUE=Liver;
RX  MEDLINE=96121368; PubMed=8575749;
RA  Kunieda T.;
RT  "Mucopolysaccharidosis type VI in rats: isolation of cDNAs encoding
RT  arylsulfatase B, chromosomal localization of the gene, and
RT  identification of the mutation.";
RL  Genomics 29:582-587(1995).
CC  -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-
CC  acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and
CC  dermatan sulfate.
CC  -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: Lysosomal.
CC  -1- DISEASE: DEFECTS IN ARSB ARE THE CAUSE OF MAROTEAUX-LAMY
CC  SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).
CC  -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
-----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
EMBL; D49434; BAA08412.1; -.
DR  HSP; P15848; IFSU
DR  InterPro: IPR000917; Sulfatase.
DR  Pfam: PF00884; Sulfatase; 1.
DR  PROSITE; PS00523; SULFATASE_1; 1.
DR  PROSITE; PS00149; SULFATASE_2; FALSE NEG.
KW  Hydrolase; Glycoprotein; Lysosome; Mucopolysaccharidosis.
FT  NON_TER 1
FT  MOD_RES 31 31 2-AMINO-3-OXOPROPIONIC ACID (BY
FT  ACT_SITE 87 87 SIMILARITY).
FT  DISULFID 57 461 POTENTIAL.
FT  DISULFID 61 95 BY SIMILARITY.
FT  DISULFID 121 132 BY SIMILARITY.
FT  DISULFID 345 387 BY SIMILARITY.
FT  CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ  SEQUENCE 473 AA; 53320 MW; 4E114C923A24AF8F CRC64;

Query Match 44.3%; Score 1333; DB 1; Length 473;
Best Local Similarity 53.9%; Pred. No. 9.6e-98;
Matches 255; Conservative 70; Mismatches 122; Indels 26; Gaps 5;

QY 94 HGSFKTPTLDKLAAGVKNLYVQICTPSRSQFITGKYIHTGLQHSIRTPQNCIL 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3 HGSVIRPHDLAAGGVLDNLYVQVLPCTPSRSOLLTGRIYHMGLOHVLMTQPCNV 62

QY 154 PLDNATLPQKLKEGVYTHMVGKHLGFYRKECMPTRRGFDFTFGSLGSDGYTH 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 63 PLDEKLLPQLLKDAGSSTHMVGKHLGMYRKECLPTRRRGFDYFYGLLGSEDTYTHEACA 122
QY 210 -YKCDSPCMCGYDLYENDNRAWDYNGIYSTQMVTORVQOQLASHNPFPFLFYAYQAV 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 PIECLNGTRCALDLRDGEEPAKEYTD-IYSTNIFTKRTATTLIANHPPKPLFLYAFOSV 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 HSPQAPGRYFEHYRSIINRRRYAAMLSCLEDAINNVTLALKTYGYNNSIIYSSDN 328
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 HDPLQVPEYMEPYDFIQKHRRYIYAGMVSLLDEAVGNVTKALKSRGLNWNVLIFSDN 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 329 GGQPTAGSNWPLRGSGTYWEGGIRAVGFVHSPLLKNKGTVCKPEVHITOMYPTLISLA 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 GGOIRSGGNWPLRGKGTWEGGIRGAGFVAPLLKQGVKSRMELMHTDPLTLVNL 301
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 389 EGOIDEDIQLDGYDIWETISEGLSRPRVDILHNTDPIY-----TRAKNGSWAAG 437
   | ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 GGSTHTGKPLDGFQVWETISEGSPRVLELLNIDPDFDGLPCPGKNTTPEKNDSPFLE 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 438 YGIWNTATQSAIRVQHWKLLTGNPGYSDWPPQSFNSLIG--PNRWNERITSSTGKSVWL 495
   : ||| : ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 HSAFNTSIHAGIRYKMKLLTGYPCGYWFPFPPSOSNISEVPS-----VDSPTKTLWL 414
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 496 FNTIADPYERVDLSNRPGIVGIVKLLRLRLSQNKTAVPVRYPPKDPSPRNLNG 548
   ||| ||| ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 FDIRNDPEERHDVSRPHPHIVQNLRLSLQYYHEHSVPSYFPFPLDPRCDPKGTG 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
ARBSE_MOUSE STANDARD;          PRT;      285 AA.
ID  ARSB_MOUSE
AC  P50429;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Arylsulfatase B (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-
DE  4-sulfatase) (G4S) (Fragments).
GN  ARSB OR AS1.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE OF 1-33 FROM N.A.
RX  MEDLINE=92241876; PubMed=1572648;
RA  Grompe M., Pieretti M., Caskey C.T., Ballabio A.;
RT  "The sulfatase gene family: cross-species PCR cloning using the MOPAC
RT  technique.";
RT  Genomics 12:755-760(1992).
RN  [2]
RP  SEQUENCE OF 34-285 FROM N.A.
RX  MEDLINE=96323200; PubMed=8710849;
RA  Evers M., Saftig P., Schmidt P., Hafner A., McLochlin D.B.,
RA  Schmah W., Hess B., von Figura K., Peters C.W.B.;
RT  "Targeted disruption of the arylsulfatase B gene results in mice
RT  resembling the phenotype of mucopolysaccharidosis VI.";
RT  Proc. Natl. Acad. Sci. U.S.A. 93:8214-8219(1996).
CC  -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-
CC  acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and
CC  dermatan sulfate.
CC  -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: Lysosomal.
CC  -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
-----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
EMBL; M82877; AAA37261.1; -.
DR  EMBL; X92096; CAA63067.1; -.

```



```
DR HSSP; P15848; IFSU.
DR MGD; MGI:88075; Asi-s.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; PARTIAL.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Hydrolase; Glycoprotein; Lysosome.
FT NON_TER 1
FT NON_CONS 33 34
FT ACT_SITE 45 45
FT CARBOHYD 86 86
FT CARBOHYD 177 177
FT CARBOHYD 189 189
FT CARBOHYD 264 264
FT NON_TER 285 285
SQ SEQUENCE 285 AA; 31727 MW; 36BA5B98EA0770C9 CRC64;

Query Match 28.7%; Score 864; DB 1; Length 285;
Best Local Similarity 49.8%; Pred. No. 5.3e-61;
Matches 164; Conservative 45; Mismatches 68; Indels 52; Gaps 3;

QY 91 VYHGSEIKPTLKLAEVGYKLENNYVQPICTPSRSQFITGKYQIHTGLOHSIIRPTQP 150
Db :|||||:|||||:|||||:|||||:
4 LGFHGSVIRTPHLDALAAGVVLNYYVQP----- 33

QY 151 NCLPLDNATLPQKLKEVGYSTHMYGKWHLGFYRKECMPTRRGFDTFGSLGSGDYVTHY 210
Db :|||||:|||||:|||||:|||||:
34 -----AGYATHMYGKWHLGMNRYKECLPTRRGFDTFYGLGSEDTYTHE 77

QY 211 KCD-----SPGMCYDLYENDNAWDYNGIYSTOMYTORVQOILASHNPTKPIFLYIAY 265
Db :|||:|||||:|||||:|||||:
78 ACAPTESLNGTRCALDLRDGEAPKEYNN-IYSTNIETKRAITVIANHPPEKPLFLYLA 136

QY 266 QAVHSPLQAPGRYEHYESIINIRRRYAAMLSCLDEAINNVTALKTYGYFNNSIIYS 325
Db :|||||:|||||:|||||:|||||:
137 QSVHDPLQPEYMEPYGFIQDKHRIYAGVSLMDAEAVGNVTALKSHGLNNVTVFIFS 196

QY 326 SDNGGQPTAGSNPLRGSKTYEGGIRAVGFVHSPILKKNKGYCKPEVHITDWTPLI 385
Db :|||||:|||||:|||||:|||||:
197 TDNGGQTRSGGNPLRGKGTLEGGIRGTGFVASPLLKQKGVKSRMLHISDLPLTV 256

QY 386 SLAGQDEIDQLDGYDIWETISEGLRSP 414
Db :|||:|||||:|||||:
257 DLGGSTGNTKPLDGFNMWKTISEGHPSP 285

RESULT 5
ARSF_HUMAN
ID ARSF_HUMAN STANDARD; PRT; 591 AA.
AC P54793;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Arylsulfatase f precursor (EC 3.1.6.-) (ASF).
GN ARSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=97336043; PubMed=9192838;
RA Puca A.A., Zollo M., Repetto M., Andolfi G., Guffanti A., Simon G.,
RA Ballabio A., Franco B.;
RT "Identification by shotgun sequencing, genomic organization, and
RT functional analysis of a fourth arylsulfatase gene (ARSF) from the
RT Xp22.3 region."
RL Genomics 42:192-199(1997).
RN [2]
RP SEQUENCE OF 370-423 FROM N.A.
RC TISSUE=Kidney;
```


RA Tomatsu S., Fukuda S., Masue M., Sukegawa K., Fukao T., Yamagishi A.,
RA Hori T., Iwata H., Ogawa T., Nakashima Y., Hanyu Y., Hashimoto T.,
RA Titani K., Oyama R., Suzuki M., Yagi K., Hayashi Y., Orii T.;
RT "Morquio disease: isolation, characterization and expression of full-
RT length cDNA for human N-acetylgalactosamine-6-sulfate sulfatase.";
RL Biochem. Biophys. Res. Commun. 181:677-683(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-95095267; PubMed-8001980;
RA Morris C.P., Guo X.H., Apostolou S., Hopwood J.J., Scott H.S.;
RT "Morquio A syndrome: cloning, sequence, and structure of the human N-
RT acetylgalactosamine 6-sulfatase (GALNS) gene.";
RL Genomics 22:652-654(1994).
RN [3]
RP VARIANT MPS-IVA LYS-204.
RX MEDLINE-92395122; PubMed-1522213;
RA Fukuda S., Tomatsu S., Masue M., Sukegawa K., Iwata H., Ogawa T.,
RA Nakashima Y., Hori T., Yamagishi A., Hanyu Y., Morooka K., Kiman T.,
RA Hashimoto T., Orii T.;
RT "Mucopolysaccharidosis type IVA. N-acetylgalactosamine-6-sulfate
RT sulfatase exonic point mutations in classical Morquio and mild
RT cases.";
RL J. Clin. Invest. 90:1049-1053(1992).
RN [4]
RP VARIANTS MPS-IVA.
RX MEDLINE-95397840; PubMed-7668283;
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Maruf Rezvi G.,
RA Yamagishi A., Yamada N., Kato Z., Isogai K., Sukegawa K., Kondo N.,
RA Suzuki Y., Shimozaawa N., Orii T.;
RT "Mucopolysaccharidosis IVA: identification of a common missense
RT mutation III13F in the N-Acetylgalactosamine-6-sulfate sulfatase
RT gene.";
RL Am. J. Hum. Genet. 57:556-563(1995).
RN [5]
RP VARIANTS MPS-IVA.
RX MEDLINE-95315929; PubMed-7795586;
RA Ogawa T., Tomatsu S., Fukuda S., Yamagishi A., Maruf Rezvi G.,
RA Sukegawa K., Kondo N., Suzuki Y., Shimozaawa N., Orii T.;
RT "Mucopolysaccharidosis IVA: screening and identification of mutations
RT of the N-acetylgalactosamine-6-sulfate sulfatase gene.";
RL Hum. Mol. Genet. 4:341-349(1995).
RN [6]
RP VARIANTS MPS-IVA ARG-77; TRP-90; VAL-96; LEU-151; GLY-230 AND THR-291.
RX MEDLINE-95359983; PubMed-7633425;
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Maruf Rezvi G.,
RA Yamagishi A., Yamada N., Kato Z., Isogai K., Sukegawa K., Kondo N.,
RA Suzuki Y., Shimozaawa N., Orii T.;
RT "Mucopolysaccharidosis type IVA: identification of six novel
RT mutations among non-Japanese patients.";
RL Hum. Mol. Genet. 4:741-743(1995).
RN [7]
RP VARIANT MPS-IVA SER-487.
RX MEDLINE-96047158; PubMed-7581409;
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Yamada N., Isogai K.,
RA Kato Z., Sukegawa K., Kondo N., Suzuki Y., Shimozaawa N., Orii T.;
RT "Two new mutations, Q473X and N487S, in a Caucasian patient with
RT mucopolysaccharidosis IVA (Morquio disease).";
RL Hum. Mutat. 6:195-196(1995).
RN [8]
RP VARIANTS MPS-IVA ALA-138; SER-151 AND LEU-151.
RX MEDLINE-96216146; PubMed-8651279;
RA Tomatsu S., Fukuda S., Yamagishi A., Cooper A., Wraith J.E., Hori T.,
RA Kato Z., Yamada N., Isogai K., Sukegawa K., Kondo N., Suzuki Y.,
RA Shimozaawa N., Orii T.;
RT "Mucopolysaccharidosis IVA: four new exonic mutations in patients
RT with N-acetylgalactosamine-6-sulfate sulfatase deficiency.";
RL Am. J. Hum. Genet. 58:950-962(1996).
RN [9]
RP VARIANTS MPS-IVA CYS-94 AND VAL-97.
RX MEDLINE-96423834; PubMed-8826435;
RA Cole D.E.C., Fukuda S., Gordon B.A., Rip J.W., Lecouteur A.N.,
RA Ruper C.A., Tomatsu S., Ogawa T., Sukegawa K., Orii T.;
RT "Heteroallelic missense mutations of the galactosamine-6-sulfate

RT sulfatase (GALNS) gene in a mild form of Morquio disease (MPS IVA).";
RL Am. J. Med. Genet. 63:558-565(1996).
RN [10]
RP VARIANTS MPS-IVA.
RX MEDLINE-97442274; PubMed-9298823;
RA Bunge S., Kleijer W.J., Tylki-Szymanska A., Steglich C., Beck M.,
RA Tomatsu S., Fukuda S., Poorthuis B.J.H.M., Czartoryska B., Orii T.,
RA Gal A.;
RT "Identification of 31 novel mutations in the N-acetylgalactosamine-6-
RT sulfatase gene reveals excessive allelic heterogeneity among patients
RT with Morquio A syndrome.";
RL Hum. Mutat. 10:223-232(1997).
RN [11]
RP VARIANTS MPS-IVA.
RX MEDLINE-98041700; PubMed-9375852;
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Ferreira P.,
RA di Natale P., Tortora P., Fujimoto A., Kato Z., Yamada N., Isogai K.,
RA Yamagishi A., Sukegawa K., Suzuki Y., Shimozaawa N., Kondo N.,
RA Sly W.S., Orii T.;
RT "Fourteen novel mucopolysaccharidosis IVA producing mutations in
RT GALNS gene.";
RL Hum. Mutat. 10:368-375(1997).
RN [12]
RP VARIANTS MPS-IVA.
RX MEDLINE-98180718; PubMed-9521421;
RA Yamada N., Fukuda S., Tomatsu S., Muller V., Hopwood J.J., Nelson J.,
RA Kato Z., Yamagishi A., Sukegawa K., Kondo N., Orii T.;
RT "Molecular heterogeneity in mucopolysaccharidosis IVA in Australia
RT and Northern Ireland: nine novel mutations including T312S, a common
RT allele that confers a mild phenotype.";
RL Hum. Mutat. 11:202-208(1998).
RN [13]
RP VARIANTS SER-393 AND MET-488.
RX MEDLINE-98112415; PubMed-9452036;
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Yamagishi A., Kato Z.,
RA Yamada N., Isogai K., Sukegawa K., Suzuki Y., Shimozaawa N., Kondo N.,
RA Orii T.;
RT "Fifteen polymorphisms in the N-acetylgalactosamine-6-sulfate
RT sulfatase (GALNS) gene: diagnostic implications in Morquio disease.";
RL Hum. Mutat. Suppl. 1:S42-S46(1998).
RN [14]
RP CATALYTIC ACTIVITY: Hydrolysis of the 6-sulfate groups of the N-
RP acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and
RP of the D-galactose 6-sulfate units of keratan sulfate.
CC [1-] SUBUNIT: OLIGOMER OF DISULFIDE LINKED 40- AND 15-KDA POLYPEPTIDES.
CC [1-] SUBCELLULAR LOCATION: Lysosomal.
CC [1-] DISEASE: Defects in GALNS are the cause of mucopolysaccharidosis
CC - type IVA (MPS-IVA) (also known as Morquio A syndrome) which is
CC characterized by specific spondyloepiphyseal dysplasia, short
CC trunk dwarfism, coxa valga, odontoid hypoplasia, corneal
CC opacities, preservation of intelligence, and excessive urinary
CC excretion of keratan sulfate and chondroitin-6-sulfate. Severely
CC affected patients usually die of cardiopulmonary disturbance or
CC cervical cord compression in the second or third decade of life.
CC [1-] SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17629; BAA04535.1;
DR EMBL; D17616; BAA04535.1; JOINED.
DR EMBL; D17617; BAA04535.1; JOINED.
DR EMBL; D17618; BAA04535.1; JOINED.
DR EMBL; D17619; BAA04535.1; JOINED.
DR EMBL; D17620; BAA04535.1; JOINED.
DR EMBL; D17621; BAA04535.1; JOINED.
DR EMBL; D17622; BAA04535.1; JOINED.
DR EMBL; D17623; BAA04535.1; JOINED.
DR EMBL; D17624; BAA04535.1; JOINED.

QY 116 YVVOPICTPSRSPITCKYQIHTGL--OHSIIIRPTQPCNCLPLDNATLPQKLKEVGYSTHM 173
 Db 109 YGSDVCTPSRSAIRGTGPIRVYGEERIFLPWTTLGLPVEVTAEMKAGAGYTTGM 168
 QY 174 VGKWLHGFYRKEC-----MPTRGEDFFGSLGSGDYTHYKCDSPGCMGYDLYENDNA 228
 Db 169 VGKWLHGINENSSDGAHLFANKGFD-FVGHNLPPFGN---SWRCDDTGL--HQDFPDTNA 222
 QY 229 AMDYDNGIYSTQMYTOR-VOQIL-----ASHNPTKPIFLYIAYQAVHSPLOAPGRYFE 280
 Db 223 CFLYNSTVAQPFQKGLQLLRDDTVGVFTEDNVNPKPFMYVFAHHHTSL-----FS 276
 QY 281 HYRSIININRRYAAMUSCLDEAINNVTLALKYGYFNNSIIYSSDNG-----GOPTA 334
 Db 277 SDFSCSTRRGYDGNLREMDQAEIQVTLVDNDIDNTVIFFTSDHGPHEVYCGE--- 333
 QY 335 GGSNWPLRSGKGYWEGGIRAVGFVHSPLLKNGKTCVKEPHVITDWPYTLISLAEGOIDE 394
 Db 334 GGDANVFRGGKSGWEGGHRIPYIVYWPGTISPQ-VSHEIVTSMDIATAVNLGGSQLPT 392
 QY 395 DIQLDGDYDIWETISEGLRSPRVDLHNDIPYTKAKNGSWAAQY 438
 Db 393 DRYDVGKCLASVLEGASSPHDFFYCKDTLMVAVRGYKKAHF 436
 RESULT 9
 STS_HUMAN STANDARD; PRT; 583 AA.
 AC P08842;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Steryl-sulfatase precursor (EC 3.1.6.2) (Steroid sulfatase) (Steryl-
 sulfate sulfohydrolase) (Arylsulfatase C) (ASC).
 GN STS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89340479; PubMed=2668275;
 RA Stein C., Hille A., Seidel J., Rijnbout S., Waheed A., Schmidt B.,
 RA Geuze H., von Figura K.;
 RT "Cloning and expression of human steroid-sulfatase. Membrane
 topology, glycosylation, and subcellular distribution in BHK-21
 cells.";
 RT J. Biol. Chem. 264:13865-13872(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87187642; PubMed=3032454;
 RA Yen P.H., Allen E., Marsh B., Mohandas T., Wang N., Taggart R.T.,
 RA Shapiro L.J.;
 RT "Cloning and expression of steroid sulfatase cDNA and the frequent
 occurrence of deletions in STS deficiency: implications for X-Y
 interchange.";
 RT Cell 49:443-454(1987).
 RN [3]
 RP SEQUENCE OF 134-174 AND 461-583 FROM N.A.
 RX MEDLINE=89077541; PubMed=3203382;
 RA Yen P.H., Marsh B., Allen E., Tsai S.P., Ellison J., Connolly L.,
 RA Neiswanger K., Shapiro L.J.;
 RT "The human X-linked steroid sulfatase gene and a Y-encoded
 pseudogene: evidence for an inversion of the Y chromosome during
 primate evolution.";
 RT Cell 55:1123-1135(1988).
 RN [4]
 RP SEQUENCE OF 22-45.
 RC TISSUE=Liver;
 RX MEDLINE=89352671; PubMed=2765556;
 RA Kawano J.-I., Kotani T., Ohtaki S., Minamino N., Matsuo H., Oinuma T.,
 RA Aikawa E.;
 RT "Characterization of rat and human steroid sulfatases.";

Biochim. Biophys. Acta 997:199-205(1989).
 [5]
 RL VARIANTS XLI LEU-341; ARG-372 AND TYR-446.
 RN MEDLINE=92170784; PubMed=1539590;
 RX Basler E., Gronpe M., Parenti G., Yates J., Ballabio A.;
 RT "Identification of point mutations in the steroid sulfatase gene of
 three patients with X-linked ichthyosis.";
 RL Am. J. Hum. Genet. 50:483-491(1992).
 CC -1- FUNCTION: CONVERSION OF SULFATED STEROID PRECURSORS TO ESTROGENS
 CC DURING PREGNANCY.
 CC -1- CATALYTIC ACTIVITY: 3-beta-hydroxyandrost-5-en-17-one 3-sulfate +
 CC H₂O = 3-beta-hydroxyandrost-5-en-17-one + sulfate.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE. THE SEQUENCE SHOWS
 CC SEVERAL MEMBRANE-SPANNING DOMAINS THAT COULD SERVE TO ANCHOR THE
 CC PROTEIN IN THE MICROSOMAL MEMBRANE.
 CC -1- DISEASE: DEFECTS IN STS ARE A CAUSE OF A VISIBLE PHENOTYPE OF
 CC SCALY SKIN, X-LINKED ICHTHYOSIS (XLI).
 CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; J04964; AAA60597.1; -;
 DR EMBL; M16505; AAA60596.1; -;
 DR EMBL; M23945; AAA60598.1; -;
 DR EMBL; M23556; AAA60599.1; -;
 DR PIR; A32641; A32641.
 DR PIR; A25961; A25961.
 DR PIR; S05415; S05415.
 DR HSP; P15848; IFSU.
 DR MTM; 308100; -;
 DR InterPro; IPR000917; Sulfatase.
 DR Pfam; PF00884; Sulfatase; 1.
 DR PROSITE; PS00523; SULFATASE_1; 1.
 DR PROSITE; PS00149; SULFATASE_2; 1.
 KW Hydrolase; Microsome; Transmembrane; Glycoprotein; Steroid metabolism;
 KW Pregnancy; Signal; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 583 STERYL-SULFATASE
 FT MOD_RES 75 75 2-AMINO-3-OXOPROPIONIC ACID (BY
 FT SIMILARITY).
 FT ACT_SITE 136 136 POTENTIAL.
 FT TRANSMEM 185 206 POTENTIAL.
 FT TRANSMEM 213 234 POTENTIAL.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 341 341 S -> L (IN XLI).
 FT /FTID=VAR_007240.
 FT VARIANT 372 372 W -> R (IN XLI).
 FT /FTID=VAR_007241.
 FT VARIANT 446 446 C -> Y (IN XLI).
 FT /FTID=VAR_007242.
 FT CONFLICT 23 23 A -> E (IN REF. 2).
 FT SEQUENCE 583 AA; 65492 MW; 74746AFA9D21A0A6 CRC64;
 SQ
 Query Match 13.7%; Score 413; DB 1; Length 583;
 Best Local Similarity 25.1%; Pred. No. 5.9e-25;
 Matches 142; Conservative 78; Mismatches 201; Indels 144; Gaps 22;
 QY 67 EPSTSTSQPHLIFLADDDGFRDVGYSR-IKTPTLDKLAAGVKL-ENYVQPICTP 124
 Db 18 EAESHAASRNILVMDLIGIDGPGCYGNKTIPTNPIDRLASGGVKLTQHLAASPLCTP 77
 QY 125 SRSQFITGKYQIHTGLQH-----SIIRPTQPCNCLPLDNATLPQKLKEVGYSTHMVGKWHL 179

[illegible]

RESULT 11

STS-RAT	STANDARD;	PRT;	577 AA.
ID AC	PI5589;		
AC AC	PI5589;		
DT DT	01-APR-1990 (Rel. 14, Created)		
DT DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE DE	Steryl-sulfatase precursor (EC 3.1.6.2) (Steroid sulfatase) (Steryl-sulfate sulfohydrolase) (Arylsulfatase C) (ASC).		
GN STS.			
OS Rattus norvegicus (Rat).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX NCBI_TaxID=10116;			
RP [1]			
RP SEQUENCE FROM N.A.			
RC STRAT=SPRAGUE-DAWLEY;			
RA Li X.M., Salido E.C., Gong Y., Yen P.H., Kitada Y., Serikawa T.,			
RA Shapiro L.J.;			
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE OF 20-45.			
RC TISSUE=Liver;			
RX MEDLINE=89352671; PubMed=2765556;			
RA Kawano J.-I., Kotani T., Ohtaki S., Minamino N., Matsuo H., Oinuma T.,			
RA Aikawa E.;			
RT "Characterization of rat and human steroid sulfatases.";			
RL Biochim. Biophys. Acta 997:199-205(1989).			
CC -1- FUNCTION: CONVERSION OF SULFATED STEROID PRECURSORS TO ESTROGENS DURING PREGNANCY.			
CC -1- CATALYTIC ACTIVITY: 3-beta-hydroxyandrost-5-en-17-one 3-sulfate + H(2)O = 3-beta-hydroxyandrost-5-en-17-one + sulfate.			
CC -1- SUBUNIT: HOMODIMER.			
CC -1- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE. THE SEQUENCE SHOWS SEVERAL MEMBRANE-SPANNING DOMAINS THAT COULD SERVE TO ANCHOR THE PROTEIN IN THE MICROSOMAL MEMBRANE.			
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.			

CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC		or send an email to license@isb-sib.ch).
CC		-----
CC	EMBL; U37138; AAC53097.1; -. .	
DR	PIR; S05414; S05414.	
DR	HSSP; P15848; IFSU.	
DR	InterPro: IPR000917; Sulfatase.	
DR	Pfam; PF00884; Sulfatase; 1.	
DR	PROSITE; PS00523; SULFATASE_1; 1.	
DR	PROSITE; PS00149; SULFATASE_2; 1.	
KW	Hydrolase; Microsome; Transmembrane; Glycoprotein; Steroid metabolism;	
KW	Pregnancy; Signal.	
FT	SIGNAL	19
FT	CHAIN	20 577
FT	MOD_RES	74 74
FT		STERYL-SULFATASE.
FT		2-AMINO-3-OXOPROPIONIC ACID (BY
FT		SIMILARITY).
FT	ACT_SITE	135 135
FT	TRANSMEM	182 202
FT	TRANSMEM	213 233
FT	CARBHYD	46 46
FT	CARBHYD	332 332
FT	CARBHYD	458 458
FT	CARBHYD	577 AA; 62678 MW; 07562EAD6DB716 CRC64;
FT	SEQUENCE	
SQ		

RESULT 12


```
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Hydrolase; Complete proteome.
FT ACT_SITE 190 190 POTENTIAL.
FT CONFLICT 191 191 M -> I (IN REF. 2).
SQ SEQUENCE 551 AA; 60717 MW; A04C8BBE5E63E988 CRC64;

Query Match 13.1%; Score 394; DB 1; Length 551;
Best Local Similarity 24.5%; Pred. No. 1.7e-23;
Matches 147; Conservative 97; Mismatches 205; Gaps 29;

QY 2 APRCAGHPPP-----PSPQACVCPGKMLAMGALAGFWILCLLTGYLSWGQALPEEEGA 57
DB 33 ARKGAGYDHPNQYLVKVPATTIADNMVPMQHPA-----QDKETQQ 73
QY 58 LLAQAGKLEPSTSTSQPHLIFILADQGRDVGYGSEI-----KPTLDKLAAGVKL 113
DB 74 KLAELKK-----TGKKPNVVFLLDVGMDVGFNGGAVAGNPTPDIDAVASQGLIL 127
QY 114 ENYVQPTSPRSQFTGKYQIHTGLQHSIIRPT---QPNCPLDNLATLPQKLEKVGYS 170
DB 128 TSAYSQSSSPTRATILITQYSH-----HGLMPPMYGQPGGLQ-GUTTLPLLHDQGYV 182
QY 171 THMYGKWHLGFYRKCMPTRRGDTF--FGSLGSGDYTHYKCDSPGCMGYDLYENDNA 228
DB 183 TOAIGKWHMG-ENKESQPNQVDFDFRGNV---SDMYTEWR-----DVHVNPEV 229
QY 229 AMDVNDGIYSTQMYTOR-----VOQILASHNP----- 255
DB 230 ALSPDRSEYIKQLPFSKDDVHVRGGEQAIADITPKYMEDLDQRMWDYGVKFLDKMAKS 289
QY 256 TKPFLFYIAQVHSPLOAQRVEHYRSI-----ININRRRYAAMLSCLEAINNVTAL 311
DB 290 DKPFLLYGTGRCH-----FDNPNKAYAGSSPARTSYGDCVMEMNDVFANLYKTL 340
QY 312 KTYGFYNNIIYSDDNGQ---PTAGSNWPLRGSGTGWEGGIRAVGFVHSPLLKNKG 368
DB 341 EKNQOLNTLIVFTSDNGPEAEVPHGRT--PFRGAGSTWEGGVVPTFYVW-----KG 393
QY 369 TVCKEP-----VHTDWPYTLISLARGIDEDTQLDGYDIWETISEGLRSPRDVILHNI 422
DB 394 MI--QPRKSDGIVDLADLFTALDLA-GH-----PGAKVANLV-----PKTTFIDGV 437
QY 423 DP----IYTRAKNGSWAGYCIWNTAQSARVQHWKLLTGNPCGYSDMV-PPQSFSLGP 477
DB 438 DQTSFFLTGNGQSNRKAHEHYFL--NGKLAAVRMDEFK-----YHVLIOQPYAYTOSGY 488
QY 478 NRWNERITSTGKSVWLFNITADPYERVDLSNRYPGIVKVKLLRRLSQFNKTAVPVRYPP 537
DB 489 QGGFTGVMTAGSSV--FNLYTPQESDSIGVRHHPGVLPQTEMHAYME--ILKYYPP 544
QY 538 K 538
DB 545 R 545

RESULT 14
STS_MOUSE STANDARD; PRT; 624 AA.
AC P50427;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Steryl-sulfatase precursor (EC 3.1.6.2) (Steroid sulfatase) (Steryl-
DE sulfatase sulfohydrolase) (Arylsulfatase C) (ASC).
GN STS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
```

Db 302 FADPGFAGRSLE-----HGAYGDSVEEMDVGVRVLAALDELGLARETLVYFTSDHCAHVEE 357
Qy 332 -----PTAGGSNWPLRSGSKGYWEGGIRAVGFVHSPILKKNKGTVCKEPVHTTDWYPTLIS 386
Db 358 LQPRGERGGSGNVFGRGKGNWEGGVPCVLRWPRELSGCRVVAEPTSLMDVFPPIVAR 417
Qy 387 LAEQSIDIDQLDGYDIWETISEGLSRSPRVDILHNIIDPIYTKAKNGSWAAGYGIWNTAIQ 446
Db 418 LAGAELPDGRVIDRDLMLPLLRGLDGAQRSEHEFLHYCNAYLQAVR--WHNG----- 466
Qy 447 SAIRVQHKLITGNPGYSDWYPPQSFSGMLGNRWHNERITSTGCKS-----VWLEFNIT 499
Db 467 SAV-----WKAFYFYP-----NFAPAGAGCGCFSTHVCILCAGPAHVTAHDPPLLEDLT 513
Qy 500 ADPYERVDLS 509
Db 514 RDPGERPLT 523
RESULT 15
ARSA_HUMAN
ID ARSA_HUMAN STANDARD; PRT; 507 AA.
AC P15289;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Arylsulfatase A precursor (EC 3.1.6.8) (ASA) (Cerebrosidase-sulfatase).
GN ARSA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90361046; PubMed=1975241;
RA Kreysing J., von Figura K., Gieselmann V.;
RT "Structure of the arylsulfatase A gene.";
RL Eur. J. Biochem. 191:627-631(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89093115; PubMed=2562955;
RA Stein C., Gieselmann V., Kreysing J., Schmidt B., Pohlmann R.,
RA Waheed A., Meyer H.E., O'Brien J.S., von Figura K.;
RT "Cloning and expression of human arylsulfatase A";
RL J. Biol. Chem. 264:1252-1259(1989).
RN [3]
RP SEQUENCE FROM N.A.
RA Adams M.D., Kerlavage A.R., Fuldner R.A., Phillips C.A., Venter J.C.;
RT "Complete sequence of a chromosome 22q subtelomeric BAC";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 19-33 AND 434-479, AND SUBUNIT5.
RX MEDLINE=92338230; PubMed=1352993;
RA Fujii T., Kobayashi T., Honke K., Gasa S., Ishikawa M., Shimizu T.,
RA Makita A.;
RT "Proteolytic processing of human lysosomal arylsulfatase A";
RL Biochim. Biophys. Acta 1122:93-98(1992).
RN [5]
RP PARTIAL SEQUENCE, AND 2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.
RX MEDLINE=95354208; PubMed=7628016;
RA Schmidt B., Selmer T., Ingendoh A., von Figura K.;
RT "A novel amino acid modification in sulfatases that is defective in
multiple sulfatase deficiency.";
RL Cell 82:271-278(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE=98191325; PubMed=9521684;
RA Lukatela G., Kraus N., Theis K., Selmer T., Gieselmann V.,
RA von Figura K., Saenger W.;
RT "Crystal structure of human arylsulfatase A: the aldehyde function
and the metal ion at the active site suggest a novel mechanism for
sulfate ester hydrolysis.";

Biochemistry 37:3654-3664(1998).
RN [7]
RP REVIEW ON MLD VARIANTS.
RX MEDLINE=95170731; PubMed=7866401;
RA Gieselmann V., Zlotogora J., Harris A., Wenger D.A., Morris C.P.;
RT "Molecular genetics of metachromatic leukodystrophy.";
RL Hum. Mutat. 4:233-242(1994).
RN [8]
RP VARIANT MLD GLN-84.
RX MEDLINE=92344341; PubMed=1353340;
RA Kappler J., von Figura K., Gieselmann V.;
RT "Late-onset metachromatic leukodystrophy: molecular pathology in two
siblings.";
RL Ann. Neurol. 31:256-261(1992).
RN [9]
RP VARIANT MLD PHE-96.
RX MEDLINE=91328147; PubMed=1678251;
RA Gieselmann V., Fluharty A.L., Tonnesen T., von Figura K.;
RT "Mutations in the arylsulfatase A pseudodeficiency allele causing
metachromatic leukodystrophy.";
RL Am. J. Hum. Genet. 49:407-413(1991).
RN [10]
RP VARIANT MLD ASP-99.
RX MEDLINE=91206410; PubMed=1673291;
RA Kondo R., Wakamatsu N., Yoshino H., Fukuhara N., Miyatake T.,
RA Tsuji S.;
RT "Identification of a mutation in the arylsulfatase A gene of a
patient with adult-type metachromatic leukodystrophy.";
RL Am. J. Hum. Genet. 48:971-978(1991).
RN [11]
RP VARIANT MLD LEU-426, AND VARIANTS CYS-193 AND SER-391.
RX MEDLINE=91074201; PubMed=1670590;
RA Polten A., Fluharty A.L., Fluharty C.B., Kappler J., von Figura K.,
RA Gieselmann V.;
RT "Molecular basis of different forms of metachromatic leukodystrophy.";
RL New Engl. J. Med. 324:18-22(1991).
RN [12]
RP VARIANT MLD SER-122.
RX MEDLINE=94063853; PubMed=7902317;
RA Honke K., Kobayashi T., Fujii T., Gasa S., Xu M., Takamaru Y.,
RA Kondo R., Tsuji S., Makita A.;
RT "An adult-type metachromatic leukodystrophy caused by substitution of
serine for glycine-122 in arylsulfatase A.";
RL Hum. Genet. 92:451-456(1993).
RN [13]
RP VARIANT MLD LEU-136.
RX MEDLINE=95163939; PubMed=7860068;
RA Kafert S., Heinisch U., Zlotogora J., Gieselmann V.;
RT "A missense mutation p136L in the arylsulfatase A gene causes
instability and loss of activity of the mutant enzyme.";
RL Hum. Genet. 95:201-204(1995).
RN [14]
RP VARIANT MLD ARG-245.
RX MEDLINE=93319632; PubMed=8101083;
RA Hasegawa Y., Kawame H., Eto Y.;
RT "Mutations in the arylsulfatase A gene of Japanese patients with
metachromatic leukodystrophy.";
RL DNA Cell Biol. 12:493-498(1993).
RN [15]
RP VARIANT MLD MET-274.
RX MEDLINE=94004907; PubMed=8104633;
RA Harvey J.S., Nelson P.V., Carey W.F., Robertson E.F., Morris C.P.;
RT "An arylsulfatase A (ARSA) missense mutation (T274M) causing late-
infantile metachromatic leukodystrophy.";
RL Hum. Mutat. 2:261-267(1993).
RN [16]
RP VARIANT MLD SER-309.
RX MEDLINE=93318834; PubMed=8101038;
RA Kreysing J., Bohne W., Bosenberg C., Marchesini S., Turpin J.C.,
RA Baumann N., von Figura K., Gieselmann V.;
RT "High residual arylsulfatase A (ARSA) activity in a patient with
late-infantile metachromatic leukodystrophy.";
RL Am. J. Hum. Genet. 53:339-346(1993).

RN [17] VARIANT SER-350.
RP MEDLINE=90083282; PubMed=2574462;
RX Gieselmann V., Polten A., Kreysing J., von Figura K.:
RA "Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal
RT and N-glycosylation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9436-9440(1989).
RN [18]
RX VARIANT MLD LEU-426.
RP MEDLINE=93202658; PubMed=8095918;
RX Barth M.L., Fensom A., Harris A.;
RA "Prevalence of common mutations in the arylsulphatase A gene in
RT metachromatic leukodystrophy patients diagnosed in Britain.";
RL Hum. Genet. 91:73-77(1993).
RN [19]
RX VARIANTS MLD VAL-212; VAL-224 AND TYR-295.
RP MEDLINE=94154687; PubMed=7906588;
RX Barth M.L., Fensom A., Harris A.;
RA "Missense mutations in the arylsulphatase A genes of metachromatic
RT leukodystrophy patients.";
RL Hum. Mol. Genet. 2:2117-2121(1993).
RN [20]
RX VARIANTS MLD LEU-82; TYR-172; CYS-201; GLN-311; VAL-335 AND TRP-390.
RP MEDLINE=96047150; PubMed=7581401;
RX Barth M.L., Fensom A., Harris A.;
RA "Identification of seven novel mutations associated with
RT metachromatic leukodystrophy.";
RL Hum. Mutat. 6:170-176(1995).
RN [21]
RX CHARACTERIZATION OF VARIANTS MET-274 AND VAL-335.
RP MEDLINE=96303701; PubMed=8723680;
RX Hess B., Kafert S., Heinisch U., Wenger D.A., Zlotogora J.,
RA Gieselmann V.;
RT "Characterization of two arylsulphatase A missense mutations D335V and
RL T274M causing late infantile metachromatic leukodystrophy.";
RN [22]
RX VARIANTS MLD.
RP MEDLINE=97245886; PubMed=9090526;
RX Draghia R., Letourneur F., Dragan C., Manicom J., Blanchot C.,
RA Kahn A., Poenaru L., Caillaud C.;
RT "Metachromatic leukodystrophy: identification of the first deletion in
RL exon 1 and of nine novel point mutations in the arylsulphatase A
gene.";
RN [23]
RX Hum. Mutat. 9:234-242(1997).
RN [24]
RX VARIANTS MLD PRO-135 AND SER-179.
RP MEDLINE=98260871; PubMed=9600244;
RX Gomez-Lira M., Perusi C., Mottes M., Pignatti P.F., Manfredi M.,
RA Rizzuto N., Salvati A.;
RT "Molecular genetic characterization of two metachromatic
RL leukodystrophy patients who carry the T799G mutation and show
different phenotypes; description of a novel null-type mutation.";
RN [25]
RX VARIANTS MLD GLN-390 AND TYR-397.
RP MEDLINE=98112481; PubMed=9452102;
RX Coulter-Mackie M.B., Gagnier L.;
RA "Two novel mutations in the arylsulphatase A gene associated with
RT juvenile (R390Q) and adult onset (H397Y) metachromatic
RL leukodystrophy.";
RN [26]
RX Hum. Mutat. Suppl. 1:S254-S256(1998).
RP VARIANT HIS-496.

RX MEDLINE=98415722; PubMed=9744473;
RA Ricketts M.H., Poretz R.D., Manowitz P.;
RT "The R496H mutation of arylsulphatase A does not cause metachromatic
Query Match 12.5%; Score 378; DB 1; Length 507;
Best Local Similarity 26.1%; Pred. No. 2.8e-22;
Matches 128; Conservative 75; Mismatches 195; Indels 92; Gaps 19;
QY 76 PHLIFILADDOGFRDVGVG-SEIKTPTLDKLAAGVKLENYVO-PICTPSRSQFTGK 133
DB 21 PNIVLIFADDLGDLGCGYHPSTTPNLDQLAAGGRFTDFYVPVSLCTPSRAALTGR 80
QY 134 YOIHGTGLQHSIIRPTQPCNLPLDNATLPQKLKEVGYSTHVMGKWHLGFYRKEC-MPTRRG 192
DB 81 LPVRMGMYPGVLVPSSRGGPLLEEVTAELAAARGYLTMAGKWHLGVGPGGAFPLPHQG 140
QY 193 FDTFFGSLGSGDYTHYK-----CDSPQCMCGYDLYENDNAAWDYDNGIYS 238
DB 141 FHRELGI-----PYSHDQGPCQLTCFPPATPCD-----DQGLVP 178
QY 239 TQM-----YTORVQOILA-SHNPTKPIFLVIAYQAVHSPLOAQGRY 278
DB 179 IPLLANLSVEAQPPLPGLLEARYMAFAHDLMAADARQDRPFELIYASHHTYP-QFSGQS 237
QY 279 FEHYRSIINRRRYAAMLSCLEAINNVTIALKTYGFYNNISIIYSSDNGQP---TAG 335
DB 238 FAE-RS---GRGPFGLDLMELDAAVGTLMTAIGDLGLEETLVIFTADNGPETMRMSRG 292
QY 336 GSNWPLRGSKTYWEGGIRAVGFVHSPLLKNKGVCKVEPVHITDWTYPTLISLAEQDIED 395
DB 293 GCSGLLRCKGKTYEGGVREPALAFWFCHTIAPG-VTHELASSLDLPTLAALAGAPL-PN 350
QY 396 IOLGDIWETIISGLRSPRVDILHNIDPIYTKAKNGSWAAGYGIWNTAIQSAIRVQHWK 455
DB 351 VTLDGFDLSPLLLTGKSPQSILF--FYPSYDVEGVFAVRTGKYKA-----H 397
QY 456 LLTGNPGYSDWVPPQSFNGLPNRWHNERITSTGKSVWLENITADPYRVDLSNRYPGI 515
DB 398 FFTQGSASHD-----TTADPACHASSSLTAH--EPELLYDLKSDPCGYNLLGGVAGA 448
QY 516 VKLLRLRLSQ 525
DB 449 TPEVLQALKQ 458
Search completed: July 23, 2002, 19:07:50
Job time: 366 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:05:29 ; Search time 356.89 Seconds
(without alignments)
10901.197 Million cell updates/sec

Title: US-09-495-823-8

Perfect score: 2266

Sequence: 1 caccgctcgccacgcgtc.....tgcacacttggtcggaattc 2266

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2234.2	98.6	2253	22	AAH46863 Human 23553 sulfat
2	2033.8	89.8	4639	22	AAAS45962 Human DNA encoding
3	2033.8	89.8	4640	21	AAC78483 Human PRO708 (UNO3
4	2033.8	89.8	4650	20	AAZ33987 Human PRO708 nucle
5	1781.6	78.6	1800	22	AAH46867 Human 23553 sulfat
6	890	39.3	1375	22	AAAS22726 Human cDNA encodin
7	736.2	32.5	1340	22	AAAS22490 Human cDNA encodin
8	625.2	27.6	729	21	AAAO2344 Human colon cancer
9	596.2	26.3	2940	22	AAH46862 Human 25278 sulfat

10	594.4	26.2	1710	22	AAH46866 Human 25278 sulfat
11	575.2	25.4	629	22	AAS41205 cDNA encoding nove
12	538.4	-23.8	562	22	AAH99356 Human protein enco
13	210.4	9.3	1615	22	AAS28640 Genomic sequence #
14	139.6	6.2	1082	21	AAAO2370 Human colon cancer
15	129	5.7	1740	23	ABLI12859 Drosophila melanog
16	123.2	5.4	294	21	AAA00278 Human colon cancer
17	110.6	4.9	1629	23	ABLI14291 Drosophila melanog
18	98.2	4.3	249	19	AAK11642 Human biallelic po
19	98.2	4.3	249	19	AAK11642 Human biallelic po
20	88.2	3.9	3039	23	ABL09675 Drosophila melanog
21	55.6	2.5	4030	23	ABLI12884 Drosophila melanog
22	55.4	2.4	4036	23	ABLI14290 Drosophila melanog
23	53	2.3	10732	21	AAAI0594 Gene encoding a su
24	52.4	2.3	2401	20	AAZ20800 Human steroid sulp
25	52	2.3	53	20	AAZ33990 Human PRO708 hybri
26	52	2.3	53	21	AAC78675 Human PRO708 hybri
27	52	2.3	11473	24	ABL33354 Human immune syste
28	51.2	2.3	1461	23	ABLI12885 Drosophila melanog
29	50.6	2.2	4006	23	ABLI12858 Drosophila melanog
30	48.2	2.1	5856	23	ABLI09674 Human immune syste
31	47.4	2.1	6090	24	ABL33578 Human immune syste
32	47	2.1	249	18	AAT89608 Hepatitis C virus
33	46.4	2.0	876	21	AAC59201 Human secreted pro
34	45.2	2.0	431	22	AAI87912 Human polynucleoti
35	45.2	2.0	3973	24	ABL32171 Human immune syste
36	45.2	2.0	16167	24	ABL33082 Human immune syste
37	45.2	2.0	16167	24	ABL34528 Human metastasis a
38	44.8	2.0	382	22	AAT80370 Human polynucleoti
39	44.8	2.0	6436	24	ABL32681 Human immune syste
40	44.6	2.0	583	22	AAAS29116 cDNA encoding for
41	44.6	2.0	700	22	AAH92784 Human inflammatory
42	44.6	2.0	700	22	AAH92785 Human inflammatory
43	44.6	2.0	4237	19	AAV61487 Human secreted pro
44	44.4	2.0	882	21	AAZ52528 Human secreted pro
45	44.4	2.0	997	22	AAC91493 Human PRO6006 cDNA

ALIGNMENTS

RESULT 1
AAH46863
ID AAH46863 standard; cDNA; 2253 BP.

XX AC AAH46863;

XX DT 25-SEP-2001 (first entry)

XX DE Human 23553 sulfatase polypeptide encoding cDNA.

XX KW Sulfatase; 23553 sulfatase; nontropic; neuroprotective; antibacterial;
antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 324..2123

XX FT FT /*tag= a

XX FT /product= "23553 sulfatase"

XX PN WO200155411-A2.

XX PD 02-AUG-2001.

XX PF 31-JAN-2001; 2001WO-US03266.

XX PR 31-JAN-2000; 2000US-0495823.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;

DR WPI; 2001-476214/51.
DR P-PSDB; AAB85483.

XX Novel human sulfatase polypeptides useful for treating and diagnosing
PT sulfatase-related disorders such as cerebrovascular diseases, acute
PT meningitis, multiple sclerosis, degenerative diseases and tumor -
XX
PS Claim 2; Fig 15; 180pp; English.

The invention provides 22438, 23553, 25278 or 26212 human sulfatase polypeptides and polynucleotides. The sulfatase genes and polypeptides are useful for treating disorders involving the brain such as cerebrovascular diseases, infections such as acute meningitis, demyelinating diseases including multiple sclerosis, degenerative diseases affecting the cerebral cortex including Alzheimer's disease and Pick disease, spinocerebellar degenerations including spinocerebellar ataxias including Friedreich ataxia, and ataxia telangiectasia, CC degenerative diseases affecting motor neurons including amyotrophic lateral sclerosis, inborn errors of metabolism such as leukodystrophies, toxic and acquired metabolic diseases, including vitamin deficiencies, CC and neurocutaneous syndromes (phakomatoses) including neurofibromatosis, CC and neurocutaneous syndromes (phakomatoses) including neurofibromatosis. CC The present sequence represents a human 23553 sulfatase polypeptide encoding cDNA.

Sequence 2253 BP; 624 A; 535 C; 554 G; 540 T; 0 other;

Query Match	98.6%	Score	2234.2	DB	22	Length	2253
Best Local Similarity	99.7%	Pred.	No. 0				
Matches 2247	Conservative	1	Mismatches	4	Indels	1	Gaps

[illegible]

Db	601	accacggatctgagattaaacacacctactctctgacaagctcgtgccgaagaggttaaac	660
Qy	661	tggagaactactatgtccagcctatttgcacaccatccagaggtcagtttattactgtgaa	720
Db	661	tggagaactactatgtccagcctatttgcacaccatccagaggtcagtttattactgtgaa	720
Qy	721	agtatcagatatcacaccgggacttcaacattctataagaacctaccacacccaactgtt	780
Db	721	agtatcagatatcacaccgggacttcaacattctataagaacctaccacacccaactgtt	780
Qy	781	tacctctggacaatgcccacctactctcagaacctgaaggaggttgattatcaacgcata	840
Db	781	tacctctggacaatgcccacctactctcagaacctgaaggaggttgattatcaacgcata	840
Qy	841	tgtctcgaaaatggcacttgggtttttacagaaaagaaatgcacgccaccagaagagat	900
Db	841	tgtctcgaaaatggcacttgggtttttacagaaaagaaatgcacgccaccagaagagat	900
Qy	901	ttgataccttttttgggttcccttttgggaagtggggattactatatacacactacaaa	960
Db	901	ttgataccttttttgggttcccttttgggaagtggggattactatatacacactacaaa	960
Qy	961	acagtctcgtggatgtgtgggctaagacttgcataagaaacgaatgtcgtcgtggactatg	1020
Db	961	acagtctcgtggatgtgtgggctaagacttgcataagaaacgaatgtcgtcgtggactatg	1020
Qy	1021	acaatggcatatactccacacagatatacacctcagagagatcacagaatcttagctccc	1080
Db	1021	acaatggcatatactccacacagatatacacctcagagagatcacagaatcttagctccc	1080
Qy	1081	ataaccaccacaagcctatatattttatatatttgcctatacaagctgttcatccaccctgc	1140
Db	1081	ataaccaccacaagcctatatattttatatatttgcctatacaagctgttcatccaccctgc	1140
Qy	1141	aagctctcgtgcaggtatttcgaacatcacagatcccatatacaacataaacggggagat	1200
Db	1141	aagctctcgtgcaggtatttcgaacatcacagatcccatatacaacataaacggggagat	1200
Qy	1201	atgctgcacatgcttctcgtctagatggaagcaatcaaacagctgacattggtctctaaaga	1260
Db	1201	atgctgcacatgcttctcgtctagatggaagcaatcaaacagctgacattggtctctaaaga	1260
Qy	1261	cttatggtttctatacaacacagattatcatcttactcttcagataatgggtgccagccta	1320
Db	1261	cttatggtttctatacaacacagattatcatcttactcttcagataatgggtgccagccta	1320
Qy	1321	cggcagagggagtaactggcctctcagaggtagcaaaagggaacatatgggaaggaggga	1380
Db	1321	cggcagagggagtaactggcctctcagaggtagcaaaagggaacatatgggaaggaggga	1380
Qy	1381	tcggggctgtaggctttgtgcataagcccactctgaaaaaacaagggaacagtgtgtaagg	1440
Db	1381	tcggggctgtaggctttgtgcataagcccactctgaaaaaacaagggaacagtgtgtaagg	1440
Qy	1441	aaactgtgcacataactgactggttaccacactctcatcttactggctgaaggacagattg	1500
Db	1441	aaactgtgcacataactgactggttaccacactctcatcttactggctgaaggacagattg	1500
Qy	1501	atgaggacattcaactagatgggctaagatatctgggagaccataagtgagggtcttcgct	1560
Db	1501	atgaggacattcaactagatgggctaagatatctgggagaccataagtgagggtcttcgct	1560
Qy	1561	caccccgagttagataatttgcataaacattgaccccatataaccaaggccaataatggct	1620
Db	1561	caccccgagttagataatttgcataaacattgaccccatataaccaaggccaataatggct	1620
Qy	1621	ccttgggcagcagcctatgggatactggaacactgcgaatccagtccagccctcagagtgcagc	1680
Db	1621	ccttgggcagcagcctatgggatactggaacactgcgaatccagtccagccctcagagtgcagc	1680
Qy	1681	actggaaattgcttacaggaaatctctgggctaagcgactgggttccccctcactgattttca	1740
Db	1681	actggaaattgcttacaggaaatctctgggctaagcgactgggttccccctcactgattttca	1740

Db	649	ggaggaggaggagaaagtgaatgtgcttggagagagagcgagccctccttgtttcttcocga	708
Qy	142	gtccatccattaaagccatcaactctctggaagattaaagtgtcgacaatggtgacagctg	201
Db	709	gtccatccattaaagccatcaactctctggaagattaaagtgtcgacaatggtgacagctg	768
Qy	202	agaggagagagattcttctgcagggtggagagtcttcaacgctctgttgggtgcattgtgt	261
Db	769	agaggagagagattcttctgcagggtggagagtcttcaacgctctgttgggtgcattgtgt	828
Qy	262	gcgcccgacscgcgcgggcgctgttctccgcgtggagatctcaactcgggacaggt	321
Db	829	gcgcccgacscgcgcgggcgctgttctccgcgtggagatctcaactcgggacaggt	888
Qy	322	gaatggctccacagggtctgtcggggacccgcctccgcctctccacagggctgtgtct	381
Db	889	gaatggctccacagggtctgtcggggacccgcctccgcctctccacagggctgtgtct	948
Qy	382	gtcctggaaagatgctagcaatggggggcgctggcaggattctgatacctctgcctcctca	441
Db	949	gtcctggaaagatgctagcaatggggggcgctggcaggattctgatacctctgcctcctca	1008
Qy	442	cttatggttacctctgtctggggcgagccttagaagaggaggaagaaggccttactag	501
Db	1009	cttatggttacctctgtctggggcgagccttagaagaggaggaagaaggccttactag	1068
Qy	502	ctcaagctggagaaaaactagagccagcacaaacttccactccagccccaactctcat	561
Db	1069	ctcaagctggagaaaaactagagccagcacaaacttccactccagccccaactctcat	1128
Qy	562	tcactctagcgatgatacagggtattagatgtgtgttaccacgagatctgagattaaaa	621
Db	1129	tcactctagcgatgatacagggtattagatgtgtgttaccacgagatctgagattaaaa	1188
Qy	622	caactactctgacaagctcgctgcggaaggagttaaactggagaactactatgtccagc	681
Db	1189	caactactctgacaagctcgctgcggaaggagttaaactggagaactactatgtccagc	1248
Qy	682	ctatttgcacaccatccaggaggtcagttattactggaaagtatcagatacacacccggac	741
Db	1249	ctatttgcacaccatccaggaggtcagttattactggaaagtatcagatacacacccggac	1308
Qy	742	tccaacattctatcataagagacctaccacccaacgaactgttaccctctggacaaatgccaccc	801
Db	1309	tccaacattctatcataagagacctaccacccaacgaactgttaccctctggacaaatgccaccc	1368
Qy	802	tacctcagaacctgaaggaggttgatattcaacgcatactgttcggaaaaatggcaacttgg	861
Db	1369	tacctcagaacctgaaggaggttgatattcaacgcatactgttcggaaaaatggcaacttgg	1428
Qy	862	gtttttacagaaaaaagtcatgccccacagaagaggtttgatacctcttttgggttccc	921
Db	1429	gtttttacagaaaaaagtcatgccccacagaagaggtttgatacctcttttgggttccc	1488
Qy	922	ttttgggaagtggggtattactatacacactcaaatgtgcagctcctcgggagtgtgtgct	981
Db	1489	ttttgggaagtggggtattactatacacactcaaatgtgcagctcctcgggagtgtgtgct	1548
Qy	982	atgacttgtataaaaacgacaatgtgctgggactatgacaactgggagatctactccaac	1041
Db	1549	atgacttgtataaaaacgacaatgtgctgggactatgacaactgggagatctactccaac	1608
Qy	1042	agatgtacactcagagatcacgcaaatcttagcttcccataaccccaaaagcctatat	1101
Db	1609	agatgtacactcagagatcacgcaaatcttagcttcccataaccccaaaagcctatat	1668
Qy	1102	ttttatatgtcctatcaagctgttcattcaccactgcaagctcctcggcaggtatttcg	1161
Db	1669	ttttatatgtcctatcaagctgttcattcaccactgcaagctcctcggcaggtatttcg	1728
Qy	1162	aacactacagatccattatacaacataaacagaggagagatatgtctgccattccttcgtct	1221
Db	1729	aacactacagatccattatacaacataaacagaggagagatatgtctgccattccttcgtct	1788

Qy	1222	tagatgaagcaatacaacaa	cgtgacatctggctctaaagac	tctatgtttctatacaaca	1281
Db	1789	tagatgaagcaatacaacaa	cgtgacatctggctctaaagac	tctatgtttctatacaaca	1848
Qy	1282	gcattatcatttactctcag	ataatggctggccagcctacg	gcagagggaggaatactggc	1341
Db	1849	gcattatcatttactctcag	ataatggctggccagcctacg	gcagagggaggaatactggc	1908
Qy	1342	ctctcagaggtagcaaaag	aacatacttggaagaggagga	tcgcggtcttaggctttgtgc	1401
Db	1909	ctctcagaggtagcaaaag	aacatacttggaagaggagga	tcgcggtcttaggctttgtgc	1968
Qy	1402	atagcccaactctgaaaaa	caagggaacagtggtgaaga	gaacctctgcacatacactgac	1461
Db	1969	atagcccaactctgaaaaa	caagggaacagtggtgaaga	gaacctctgcacatacactgac	2028
Qy	1462	ggtaccccaactctcattt	caactggcttgaaggacagat	tgatgaggacattcaactagatg	1521
Db	2029	ggtaccccaactctcattt	caactggcttgaaggacagat	tgatgaggacattcaactagatg	2088
Qy	1522	gctatgatatctgggagac	caataagtggaggttcttgc	tcccccagctagatatatttgc	1581
Db	2089	gctatgatatctgggagac	caataagtggaggttcttgc	tcccccagctagatatatttgc	2148
Qy	1582	ataacattgaccccatata	caccaagggcaaaaaatgg	ctctggcgacagagctatggga	1641
Db	2149	ataacattgaccccatata	caccaagggcaaaaaatgg	ctctggcgacagagctatggga	2207
Qy	1642	tctggaaactgcaatccag	ctcagccatcagagtgacag	ctctggaacttgcttcacagaa	1701
Db	2208	tctggaaactgcaatccag	ctcagccatcagagtgacag	ctctggaacttgcttcacagaa	2267
Qy	1702	atcttggtcagcgacgtgg	gtccccctcagctcttccag	caacctgggacccagcccggt	1761
Db	2268	atcttggtcagcgacgtgg	gtccccctcagctcttccag	caacctgggacccagcccggt	2327
Qy	1762	ggcaaatgaacggatcac	ctctgcacactggcaaaagt	gtatggcttttcaacatcacag	1821
Db	2328	ggcaaatgaacggatcac	ctctgcacactggcaaaagt	gtatggcttttcaacatcacag	2387
Qy	1822	ccgaccataatgagaggt	tggaacctatctaaacaggt	atccagaatactggaagaagctcc	1881
Db	2388	ccgaccataatgagaggt	tggaacctatctaaacaggt	atccagaatactggaagaagctcc	2447
Qy	1882	tacggaggtctcacagtt	caacaaactgcagtgccgt	tcaggtatcccccaagacc	1941
Db	2448	tacggaggtctcacagtt	caacaaactgcagtgccgt	tcaggtatcccccaagacc	2507
Qy	1942	ccagaagtcaaccttagct	caatggaggggtcttaggac	catggtatgagagagaaccca	2001
Db	2508	ccagaagtcaaccttagct	caatggaggggtcttaggac	catggtatgagagagaaccca	2567
Qy	2002	agaaaaagaaagcaagca	aaaaatcaggcttgagaaaaa	gcaaaagaaaaagc-aaaaaaga	2060
Db	2568	agaaaaagaaagcaagca	aaaaatcaggcttgagaaaaa	gcaaaagaaaaagc-aaaaaaga	2627
Qy	2061	agaagaacagcagaagc	agctctcaggttcaactgc	cattcaggt 2107	
Db	2628	agaagaacagcagaagc	agctctcaggttcaactgc	cattcaggt 2674	

```

.RESULT      3
AAC78483
ID AAC78483 standard; cDNA; 4640 BP.
XX
AC AAC78483;
XX
XX 08-FEB-2001 (first entry)
XX
DE Human PRO708 (UNQ372) nucleotide sequence SEQ ID NO:113.
XX
XX Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW

```

Human; secreted protein; transmembrane protein; PRO; EST; cytosstatic;

Db 1909 cctcagaggtagcaaggaacataattgggaaggaggtacccggctgtaggctttgtgc 1968
Qy 1402 atagcccacttctgaaacaagggaacagtggttaaggaaacctgtcacatcactgact 1461
Db 1969 atagcccacttctgaaacaagggaacagtggttaaggaaacctgtcacatcactgact 2028
Qy 1462 ggtaccacactctcatttccactggctggaagacagatgatgagacattcaactagatg 1521
Db 2029 ggtaccacactctcatttccactggctggaagacagatgatgagacattcaactagatg 2088
Qy 1522 gctatgatcttggagaccataaagtgggtcttcctgcctcaaccccgagtagattttgc 1581
Db 2089 gctatgatcttggagaccataaagtgggtcttcctgcctcaaccccgagtagattttgc 2148
Qy 1582 ataacattgacccctatcaccaaggcaaaaaatggtctctggcagcaggtatgga 1641
Db 2149 ataacattgacccc-tatcaccaaggcaaaaaatggtctctggcagcaggtatgga 2207
Qy 1642 tctggaacactgcaatccagtcagccatcagatgcagcactggaattgcttacaggaa 1701
Db 2208 tctggaacactgcaatccagtcagccatcagatgcagcactggaattgcttacaggaa 2267
Qy 1702 atcctggctacagcagctgggttccccctcagctcttccagcaacctgggacgacccggt 1761
Db 2268 atcctggctacagcagctgggttccccctcagctcttccagcaacctgggacgacccggt 2327
Qy 1762 ggcacaaatgaacggtatccctgcctcaactgcgaagtgatggtctttcaacatcacag 1821
Db 2328 ggcacaaatgaacggtatccctgcctcaactgcgaagtgatggtctttcaacatcacag 2387
Qy 1822 cgcacccatagagaggtggacctatctaacaggtatccaggaatcgtgaagaagctcc 1881
Db 2388 cgcacccatagagaggtggacctatctaacaggtatccaggaatcgtgaagaagctcc 2447
Qy 1882 tacgagagctctcacagttcaacaaaatgcagtgccggtcaggtatcccccccaagacc 1941
Db 2448 tacgagagctctcacagttcaacaaaatgcagtgccggtcaggtatcccccccaagacc 2507
Qy 1942 cagaagtaaacctagctcaatgaggggtcttagggaccatggtatagagaggaacca 2001
Db 2508 cagaagtaaacctagctcaatgaggggtctgaggaccatggtatagagaggaacca 2567
Qy 2002 agaaaaaagaacgaacaaatcaggctgagaaaaagcaaaagaaaaagc-aaaaaaaga 2060
Db 2568 agaaaaaagaacgaacaaatcaggctgagaaaaagcaaaagaaaaagcaaaagaaaga 2627
Qy 2061 agaaaaaagaacgaacaaatcaggctgagaaaaagcaaaagaaaaagcaaaagaaaga 2107
Db 2628 agaaaaaagaacgaacaaatcaggctgagaaaaagcaaaagaaaaagcaaaagaaaga 2674

RESULT 5
AAH46867
ID AAH46867 standard; cDNA; 1800 BP.
AC AAH46867;
XX
XX
XX
XX

DT 25-SEP-2001 (first entry)
XX
DE Human 23553 sulfatase polypeptide coding sequence.
XX

KW Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;
KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.
XX

OS Homo sapiens.

XX Key Location/Qualifiers
FH 1..1800
CDS /*tag= a
FT /product= "23553 sulfatase"
XX

PN W0200155411-A2.

XX 02-AUG-2001.
XX
XX 31-JAN-2001; 2001WO-US03266.
XX
XX 31-JAN-2000; 2000US-0495823.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
XX
XX WPI; 2001-476214/51.
DR P-PSDB; AAB85483.
XX
XX Novel human sulfatase polypeptides useful for treating and diagnosing
PT sulfatase-related disorders such as cerebrovascular diseases, acute
PT meningitis, multiple sclerosis, degenerative diseases and tumor -
XX
XX Claim 2; Fig 15; 180pp; English.
XX
XX The invention provides 22438, 23553, 25278 or 26212 human sulfatase
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides
CC are useful for treating disorders involving the brain such as
CC cerebrovascular diseases, infections such as acute meningitis,
CC demyelinating diseases including multiple sclerosis, degenerative
CC diseases affecting the cerebral cortex including Alzheimer's disease
CC and Pick disease, spinocerebellar degenerations including spinocerebellar
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,
CC degenerative diseases affecting motor neurons including amyotrophic
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,
CC toxic and acquired metabolic diseases, including vitamin deficiencies,
CC and neurocutaneous syndromes (phakamatoses) including neurofibromatosis.
CC The present sequence represents a human 23553 sulfatase polypeptide
CC coding sequence.
XX
XX Sequence 1800 BP; 534 A; 434 C; 423 G; 409 T; 0 other;

Query Match 78.6%; Score 1781.6; DB 22; Length 1800;
Best Local Similarity 99.7%; Pred No. 0;
Matches 1795; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
Qy 324 atggctccaggggctgtgctgggggcatccgctccgccttccacagcctgtgtctgt 383
Db 1 atggctccaggggctgtgctgggggcatccgctccgccttccacagcctgtgtctgt 60
Qy 384 cctggaagagatgctagcaatggggcgctggcagattctggtatcctcgcctcctcaat 443
Db 61 cctggaagagatgctagcaatggggcgctggcagattctggtatcctcgcctcctcaat 120
Qy 444 tatggttacctgtcctggggcaggcccttagaaggaggaggaagggccttactagct 503
Db 121 tatggttacctgtcctggggcaggcccttagaaggaggaggaagggccttactagct 180
Qy 504 caagctggagagaaactagagcccgagcaacttccactccacgcccactatcttttc 563
Db 181 caagctggagagaaactagagcccgagcaacttccactccacgcccactatcttttc 240
Qy 564 atcctagcggatgatcaggagatttagagatgtgggttaccacgagatctgagattaaaaa 623
Db 241 atcctagcggatgatcaggagatttagagatgtgggttaccacgagatctgagattaaaaa 300
Qy 624 cctactcttgacaagctcgctccgaagagagtgtaactggagaactactatgtccagcct 683
Db 301 cctactcttgacaagctcgctccgaagagagtgtaactggagaactactatgtccagcct 360
Qy 684 atttgacacaccatccaggagtcagttttattactggaagatcatagatacacacccgactt 743
Db 361 atttgacacaccatccaggagtcagttttattactggaagatcatagatacacacccgactt 420
Qy 744 caacattctatcataagacacctaccaccccaacttttacctctgtgacaatgccacctta 803
Db 421 caacattctatcataagacacctaccaccccaacttttacctctgtgacaatgccacctta 480

Qy	1884	cggaggctctcacaggttccaacaaactgcagtgccggtcaggtatcccccaaaagacccc	1943
Db	1561	cggaggctctcacaggttccaacaaactgcagtgccggtcaggtatcccccaaaagacccc	1620
Qy	1944	agaagtaaccctaggctcaatgcaggggtctagggaccatgatatagagagaaacccaag	2003
Db	1621	agaagtaaccctaggctcaatgcaggggtctagggaccatgatatagagagaaacccaag	1680
Qy	2004	aaaaagaagccaagcgaataatcaggctgcagaaagcaaaaagcaaaaagc-aaaaaaaggaag	2062
Db	1681	aaaaagaagccaagcgaataatcaggctgcagaaagcaaaaagcaaaaagcaaaaaggaag	1740
Qy	2053	aagaaacagcagaaagcagtcctcaggttccaacttgcattcagggttacttctgtgataa	2122
Db	1741	aagaaacagcagaaagcagtcctcaggttccaacttgcattcagggttacttctgtgataa	1800
RESULT	6		
ID	AAS22726	standard; cDNA; 1375 BP.	
XX	AC	AAS22726;	
XX	DT	24-OCT-2001 (first entry)	
XX	DE	Human cDNA encoding a novel human protein #292.	
XX	KW	Human; novel protein; ss; Antianemic; osteopathic; antiinflammatory; immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic; anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral; antitubercial; antiallergic; dermatological; haemostatic; antiasthmatic; thrombolytic; immunogen; antibody; gene therapy; neurological disorder; Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis tissue regeneration; immune disorder.	
OS	XX	Homo sapiens.	
PN	XX	WO200155437-A2.	
PD	XX	02-AUG-2001.	
PF	XX	25-JAN-2001; 2001WO-US02623.	
PR	XX	25-JAN-2000; 2000US-0491404.	
PA	XX	(HYSE-) HYSEQ INC.	
PI	XX	Tang YT, Liu C, Drmanac RT;	
DR	XX	WPI; 2001-451939/48.	
DR	XX	P-PSDB; AAU14421.	
PT	XX	Isolated polypeptides useful for treating anti-inflammatory diseases, nervous system disorders, and for regenerating bone and cartilage -	
PS	XX	Claim 1; Page 699; 894pp; English.	
CC	XX	The invention relates to polynucleotides encoding novel human proteins or their active domains. The polypeptides, polynucleotides and antibodies raised against the polypeptides are used in a method of treatment of a mammal and prevention of disorders caused by the aberrant protein expression or activity. The polypeptides can be used as molecular weight markers, food supplements, and in antibody production. The polypeptides are used to identify compounds which bind to the polypeptides. Polynucleotides of the invention are used as probes and primers, for sequencing, for chromosome or gene mapping, in the production of recombinant proteins, and in generating anti-sense DNA or RNA and in gene therapy. Polypeptides of the invention can be used to target drugs to a tumour, in assays to determine biological activity, to raise antibodies/elicit an immune response, to determine quantitative protein levels, as tissue markers, and to isolate receptors or ligands. Polypeptides of the invention may also be useful in treating platelet disorders, stem cell disorders, regenerating bone, cartilage, tendon,	

Query Match 32.5%; Score 736.2; DB 22; Length 1340;
Best Local Similarity 94.3%; Pred. No. 8e-201;
Matches 796; Conservative 0; Mismatches 44; Indels 4; Gaps 3;

Qy 369 caggcctgtctctcgtgaaagatgtagcaatggtggcgctggcagattctgcatc 428
Db 484 caggcctgtctctcgtgaaagatgtagcaatggtggcgctggcagattctgcatc 543

Qy 429 ctctgctcctcactatgttacctgctcctggcgccagccttagaagaggagaa 488
Db 544 ctctgctcctcactatgttacctgctcctggcgccagccttagaagaggagaa 603

Qy 489 gggccttactagctcaagctggagagaaactagagccagcaacttccacctccag 548
Db 604 gggccttactagctcaagctggagagaaactagagccagcaacttccacctccag 663

Qy 549 cccatctcatttccctcagcagatgtagcagggatttagagatgtgggttacacgga 608
Db 664 cccatctcatttccctcagcagatgtagcagggatttagagatgtgggttacacgga 723

Qy 609 tctgagattaaacacactactcttgacaagctcgtccgaaaggagttaaactggaac 668
Db 724 tctgagattaaacacactactcttgacaagctcgtccgaaaggagttaaactggaac 783

Qy 669 tactatgtccagcctatttgcacaccatccaggagtgagtttatttactggaagtacag 728
Db 784 tactatgtccagcctatttgcacaccatccaggagtgagtttatttactggaagtacag 843

Qy 729 atacacccggacttcaacattcttatataagacctaccacccaaactgtttacctctg 788
Db 844 atacacccggacttcaacattcttatataagacctaccacccaaactgtttacctctg 903

Qy 789 gacaatccacccctacctcagaaactgaagaggttgatattcaacgatgtgcgga 848
Db 904 gacaatccacccctacctcagaaactgaagaggttgatattcaacgatgtgcgga 963

Qy 849 aaatggcacttgggtttttacagaaaagaaatgcagtcgccaccagaaaggatttgatacc 908
Db 964 aaatggcacttgggtttttacagaaaagaaatgcagtcgccaccagaaaggatttgatacc 1023

Qy 909 ttttttgggtcccttttgggaagtggggattactatatacacactacaaatgtgacagt-cc 967
Db 1024 ttttttgggtcccttttgggaagtggggattactatatacacactacaaatgtgacagtcc 1083

Qy 968 tgggatgtggctatgactgttatgaaacgacaatgctgcctgggaactatgacaatgg 1027
Db 1084 tgggatgtggctatgactgttatgaaacgacaatgctgcctgggaactatgacaatgg 1143

Qy 1028 catacttccacacagatgtacactcagagatgacagcaaatcttagcttccataacc 1087
Db 1144 catacttccacacagatgtacactcagagatgacagcaaatcttagcttccataacc 1203

Qy 1088 cacaagcctatatttttatatatgtcctatcaagctgttcaatccacactgcaagct-c 1146
Db 1204 cacaagcctatatttttatatatgtcctatcaagctgttcaatccacactgcaagctcc 1263

Qy 1147 ctggcaggtatttcgaacact--accgatccattatcaacataaacagaggagatatgc 1204
Db 1264 ctggcaggtatttcgaacacttaccggcccttattcaacataaacagaggaggattg 1323

Qy 1205 tgcc 1208
Db 1324 gccc 1327

RESULT 8
AAA02344
ID AAA02344 standard; cDNA; 729 BP.
XX
AC AAA02344;
XX
DT 19-MAY-2000 (first entry)

XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2335.
DE
XX
KW Human: colon cancer; tumour; diagnosis; gene expression product;
KW probe: detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN W09958675-A2.
XX
PD 18-NOV-1999.
XX
XX 13-MAY-1999; 99WO-US10602.
XX
XX 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
XX WPI; 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
PT cells -
XX
XX Claim 1; Page 924-925; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotide sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
XX Sequence 729 BP; 232 A; 178 C; 163 G; 143 T; 13 other;

Query Match 27.6%; Score 625.2; DB 21; Length 729;
Best Local Similarity 97.3%; Pred. No. 4.4e-169;
Matches 644; Conservative 0; Mismatches 17; Indels 1; Gaps 1;

Qy 1407 ccacttctgaaaaaagggaacagatgtgtaagaaacctgtgcacatcactgactggtac 1466
Db 9 cccctctttnnaaacagggaacagatgtgtaagaaacctgtgcacatcactgactggtac 68

Qy 1467 cccactctcatttcactggtggaaggacagattgatgaggacattcaactagatggtcat 1526
Db 69 cccactctcatttcactggtggaaggacagattgatgaggacattcaactagatggtcat 128

Qy 1527 gatattctgggagaccataaagtggaggttcttcgctcaccctccagtagatttttgataac 1586
Db 129 gatattctgggagaccataaagtggaggttcttcgctcaccctccagtagatttttgataac 188

Db 977 ggcgcagccatctctctgcccagccagccctcagcgtcccctctctctctctatgtgccc 1036
Qy 1117 atcaagctgttcattccaccactgcaagctcctggcaggtatttcgaaacactaccgattcca 1176
Db 1037 tccaggcagttacacacacccctcagtcctcctgtgagtacctgtaccgtctaccgaccca 1096
Qy 1177 ttataacataaaacagagagagatgctgccatgcttctctctgttagatgaagaatcaa 1236
Db 1097 tgggcgaatgtggccggcggaagtcacgcggcctgtgtgacctgcatggtgagctgtgc 1156
Qy 1237 acaagctgacattgctctaaagacttatggtttctataacaaacagcattatcttact 1296
Db 1157 gcaacataccctggccctcaagcgcgtacggtttctcaacaaacagtgcatctctct 1216
Qy 1297 cttcagataatgtgtggccagcctcagcgagggaggttaactggcctctcagaggtagca 1356
Db 1217 ccagtgaacatgtgtggccagactttctcggggcgacgaactggcgtcccgaggaagca 1276
Qy 1357 aaggaacatatgtggaaggggggtccgggctgtaggtctgtgtgcatagccactttctga 1416
Db 1277 agggcacttattgggaagtgcggtggcgccctaggtgttctcaacaaacagtgccctgtca 1336
Qy 1417 aaaacaagggaacagtggttaaggaaacctgtgcacatacactgactggttaccocactctca 1476
Db 1337 agcgaagcaacgacagccggcactgattgcacatacactgactggttaccocactctgg 1396
Qy 1477 ttccactgctgaaggacagattgatgaggaattcaactatgactgactggttaccocactctgg 1536
Db 1397 tgggtctggcaggtgtgtaccactcagcagccggtggtggtgactggttaccocactctgg 1456
Qy 1537 agaccataagtgaggtgtcttcgaccccgagtagatatattttgcataaacttgaccacca 1596
Db 1457 cggccatcagcgagggcgccggtccacacgcacgagatcctgcacaaacttgaccacca 1516
Qy 1597 tatacacaaggcaaaaatgctcctggcgagcaggtatgggtatctggaacactgcaa 1656
Db 1517 tctacaacctgcccagatggctccctggagggcggttggcatctggaacaccgcg 1576
Qy 1657 tccagtcagccatcagatgcagcactggaattgcttacaggaatactctggtcacagcg 1716
Db 1577 tgcaggtgcatacccgctgggtgagtggaagctgctgacaggaagaccocgctatggcg 1636
Qy 1717 actgggttccccctcagcttcttcagcaacctggtgacogaaccggtggcacaatgaacgga 1776
Db 1637 attgatccaccgcagacactggccaccttcccgggtgagctggtggaacctggaaacgaa 1696
Qy 1777 tcacctgtaactggcaaaagtgtatgcttttcaacatacagcgcagccactatgaga 1836
Db 1697 tggcc---agtgtccgcagggcggtgtggctctcaacatacagtgctgaccttatgaac 1793
Qy 1837 ggggtgacctatctaaacaggtatccaggaatcgtgaagaagctctcactcagggtctcac 1896
Db 1754 gggaggacctggtggccagcgccctgatgtggtccgcacctgctggtcgcctggcg 1813
Qy 1897 agttcaacaacactcagtcagtcggttcaggtatcccccaagaccocccagaagtaacccta 1956
Db 1814 aataaaccgcacagccactcccgtaagctaccacagctgagaaccccggtctatctg 1873
Qy 1957 ggcctcaatggaggggtctagggaacatggtatagagagaacaaagaaagaaagcaaa 2016
Db 1874 acttttaa tgggggtgtcttgggggcccgtggccagtgatgaggaagagaggaaggaag 1933
Qy 2017 gca 2019
Db 1934 gga 1936

RESULT 10
AAH46866
ID AAH46866 standard; cDNA: 1710 BP.
XX
AC AAH46866;

XX 25-SEP-2001 (first entry)
XX Human 25278 sulfatase polypeptide coding sequence.
DE
XX
XX Sulfatase; 25278 sulfatase; nootropic; neuroprotective; antibacterial;
KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.
XX
XX Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1710
FT /tag= a
FT /product= "25278 sulfatase"
XX
XX WO200155411-A2.
XX
XX 02-AUG-2001.
XX
XX 31-JAN-2001; 2001WO-US03266.
XX
XX 31-JAN-2000; 2000US-0495823.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
PI
XX
XX WPI: 2001-476214/51.
XX
XX P-PSDB; AAB85482.
XX
XX Novel human sulfatase polypeptides useful for treating and diagnosing
PT sulfatase-related disorders such as cerebrovascular diseases, acute
PT meningitis, multiple sclerosis, degenerative diseases and tumor -
XX
XX Claim 2; Fig 10; 180pp; English.
XX
XX The invention provides 22438, 23553, 25278 or 26212 human sulfatase
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides
CC are useful for treating disorders involving the brain such as
CC cerebrovascular diseases, infections such as acute meningitis,
CC demyelinating diseases including multiple sclerosis, degenerative
CC diseases affecting the cerebral cortex including Alzheimer's disease
CC and Pick disease, spinocerebellar degenerations including spinocerebellar
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,
CC degenerative diseases affecting motor neurons including amyotrophic
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,
CC toxic and acquired metabolic diseases, including vitamin deficiencies,
CC and neurocutaneous syndromes (phakmatoses) including neurofibromatosis.
CC The present sequence represents a human 25278 sulfatase polypeptide
XX coding sequence.
XX
XX Sequence 1710 BP; 340 A; 554 C; 500 G; 316 T; 0 other;

Query Match 26.2%; Score 594.4; DB 22; Length 1710;
Best Local Similarity 62.2%; Pred. No. 5.2e-160;
Matches 1007; Conservative 0; Mismatches 596; Indels 15; Gaps 4;

Qy 402 atgagggcgctggcagattcttgatcctctgctccctcacttatggttacctgctctgg 461
Db 1 atgcacacccctcaactgcttctctctggtcagcgtctgcagcttctgctacctgctctgg 60
Qy 462 ggcagggccttagaagagaggaaggggccttactactgaagctggagagaaacta 521
Db 61 gactggggcc-----aagcagagcttcgtggcgcagcggcccgaggctggcga--g 111
Qy 522 gagccagacacaattccacctccagcccatctctatcttctcactgagcagatgacag 581
Db 112 cagccctggcgtctccgcgccagcctccacacatcttctcactcactcagcagaccaa 171
Qy 582 ggatttagagatgggtttaccacggatctgagattaaacacacttctcttgacaagctc 641
Db 172 ggcatacacgacgtggggtcaccatgggttcagatatcgagacccctacgtgacggctg 231

Qy	642	gctgcgaaagagctaaactcgagaaactactatgtccagccctatttgcacacatccagg	701
Db	232	gcgcgaaggggctcaagttggagaattattacatccagccactctgcgccttcgcgg	291
Qy	702	agtcagttattactggaagatcacagatacacacccggaactcaacatttcatataaga	761
Db	292	agcagctctactctgcggatccagatccacagagatcccgatctccatccatccgc	351
Qy	762	ctcacccaaaccacactgtttacctctggacaatgccacctactctcagaactgaagag	821
Db	352	ccacagcagcccaactgctgcctctggaccagtgcaactgccacagagctgcagag	411
Qy	822	gttggatatccaacgcataatgtctcgsgaaatggcaacttgggtttttacagaaagaatgc	881
Db	412	gcaggttattccaccatagtgtggcgaagtgcgacctggctctaccggaaggagtgt	471
Qy	882	atgccacagaaagagatttgataccttttttgttcccttttgggaagtggggattac	941
Db	472	ctgcccaccctcggggcttcgacaccttctcggtctcgctccagcgcaatgtggacctat	531
Qy	942	tatacactacaaatgtgacagctcctggatgtgtggtctatgacttgtatgaaaacgac	1001
Db	532	tacacctatgaacactgtgatggccacggctgtgcgcttcgacctgcacgaggtgcag	591
Qy	1002	aatgtcctctggactatgacaatggcatactaccacacagatgtacactcagagagta	1061
Db	592	aatgtggcttggg---ggctcagcgccagttactccactatgtcttaagccccagcgcc	648
Qy	1062	cagcaaatcttaagcttcccatcaaccccaagaagctatattttatatattgctctcaaa	1121
Db	649	agccatactctgcgcagcaacagccctcaagctccccctctctctatgtggccctccag	708
Qy	1122	gctgttcaattcaaccactgcaagctcctgcgaggtatttcgcaactaccgcatccatttc	1181
Db	709	gcagttacacacccctgcagtcctcctcgtgagtacctgttacgctaccgcaccatggcc	768
Qy	1182	aacataaacagagagagatatgctgccatgctttcctgcttagatgaagcaatacaaac	1241
Db	769	aatgtggccggcgggaagtcacgcggccatggtgcacctgcatggatgaggtgtgcgaac	828
Qy	1242	gtgacattggctctaaagactatggtttctatacaacacagcatattacttacttca	1301
Db	829	atcacctggccctcaagcgtacggtttctacaacaacagtgctatcatctctccagt	888
Qy	1302	gataatggtggccagcctacgcgagggaggtaactggctcctcagaggtagcaaaagga	1361
Db	889	gacaaatgtggccagacttcttcggggggcagcaactggcgctccgaggaacgaaggcc	948
Qy	1362	acataattggaaaggagatccggctgtaggcttttgcatagcccacttctgaaaaac	1421
Db	949	acttattgggaagtggcgtgcggggcttaggctttgtcccacgtcccctgtcgaacgga	1008
Qy	1422	aagggaacagtgtgtaaggaaacctgtgcatactactggtgtaccaccaactctattca	1481
Db	1009	aagcaacgacaaacggcgactgatgcacatccactgactgtaccgcacctgtgtgggt	1068
Qy	1482	ctggctgaaggacagattgatagagacatccaactagatggtctatgatcttgggagacc	1541
Db	1069	ctggcaggtgtgtaccacctcagacgcgatgggtcgttagtggctacgactgtgtgcggcc	1128
Qy	1542	ataagtgggggtcttcgctcaaccgagtagatatattgcataacattgaccccatatc	1601
Db	1129	atcagcgagggccgggcttcaacgcgcagcgagacatcccgcaacattgacccactatc	1188
Qy	1602	accaagcgaaaaaatggctcctggcgacgaggtgatggatctgtggaactgtgcaatccag	1661
Db	1189	aaccaatgccagatggctccctcgaggcggttgcactgtgcatacgcgcgtgcag	1248
Qy	1662	tcagccaatcagatgcagcactggaaaattgtctacagaaaatctctgctacagcactgg	1721
Db	1249	gctgcataccgggtgggtgagtggaagctgctgcagagagaccgcgctatgtggcattgg	1308

Qy	1722	gtccccctcagtcctttccaggacaacctgggacgcgaaccgggtgccacaatgaacgggatcaacc	1781
Db	1309	attcccaccgcadacactggccaaccttcccggttagctggtgaaacctggaaagaatggccc	1368
Qy	1782	tctgcaactggcaaaagtgtatggcttttcaaaccatacacagccgacccatatgagaggggtg	1841
Db	1369	---agtctccgcacggcgctggtgctcttcaacatacagctgcgaccttatgaacgggag	1425
Qy	1842	gacctattcaacagtgatccacaggaatctgtgaagaagctcctacgaggtctctcacagttc	1901
Db	1426	gacctggtggtccagcgccctgatgtggtccgacccctgctgctgctgctgctgctgctgctg	1485
Qy	1902	aacaaaaactcagtcgccgtaagttatcccccaaacaccaccaagaagtaaaccctaggctc	1961
Db	1486	aaccgcacagcatccccgttcgctaccagctggaaccccccggtcctcatctgacttt	1545
Qy	1962	aatgaggggtctagggaaccatggtatagagaggaaaccaaagaaaagagccaagca	2019
Db	1546	aatgggggtgcttggggccctggccagtgatgaggaaggaggaggaaggaggga	1603
RESULT_11			
AA	AAS41205	ID AAS41205 standard; cDNA; 629 BP.	
AC	AAS41205;		
XX	17-DEC-2001 (first entry)	cDNA encoding novel human enzyme polypeptide #421.	
DE		Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase	
KW		ligase; hyperproliferative disorder; immunodeficiency disorder;	
KW		autoimmune disorder; neurological disorder; metabolic disorder;	
KW		inflammatory disorder; cardiovascular disorder; reproductive disorder;	
KW		blood-related disorder; infectious disorder; gene therapy; cytostatic;	
KW		anti arthritic; nephrotropic; anticoagulant; ss.	
OS	Homo sapiens.		
XX	WO200155301-A2.		
PN	02-AUG-2001.		
PD			
PF	17-JAN-2001; 2001WO-US01239.		
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		

PR	14-AUG-2000;	2000US-02235750;	2000US-02235751;
PR	14-AUG-2000;	2000US-02235752;	2000US-02235753;
PR	14-AUG-2000;	2000US-02235759;	2000US-02235759;
PR	18-AUG-2000;	2000US-02262729;	2000US-02262729;
PR	22-AUG-2000;	2000US-02266811;	2000US-02266811;
PR	22-AUG-2000;	2000US-02266868;	2000US-02266868;
PR	22-AUG-2000;	2000US-02271882;	2000US-02271882;
PR	23-AUG-2000;	2000US-02277009;	2000US-02277009;
PR	30-AUG-2000;	2000US-02289824;	2000US-02289824;
PR	01-SEP-2000;	2000US-02232877;	2000US-02232877;
PR	01-SEP-2000;	2000US-02234343;	2000US-02234343;
PR	01-SEP-2000;	2000US-02234344;	2000US-02234344;
PR	01-SEP-2000;	2000US-02233445;	2000US-02233445;
PR	05-SEP-2000;	2000US-02295913;	2000US-02295913;
PR	05-SEP-2000;	2000US-02230437;	2000US-02230437;
PR	06-SEP-2000;	2000US-02304338;	2000US-02304338;
PR	08-SEP-2000;	2000US-02312423;	2000US-02312423;
PR	08-SEP-2000;	2000US-02312424;	2000US-02312424;
PR	08-SEP-2000;	2000US-02312424;	2000US-02312424;
PR	08-SEP-2000;	2000US-02311413;	2000US-02311413;
PR	08-SEP-2000;	2000US-02314114;	2000US-02314114;
PR	08-SEP-2000;	2000US-02320800;	2000US-02320800;
PR	08-SEP-2000;	2000US-02320811;	2000US-02320811;
PR	12-SEP-2000;	2000US-02319681;	2000US-02319681;
PR	14-SEP-2000;	2000US-02323397;	2000US-02323397;
PR	14-SEP-2000;	2000US-02323398;	2000US-02323398;
PR	14-SEP-2000;	2000US-02323399;	2000US-02323399;
PR	14-SEP-2000;	2000US-02324000;	2000US-02324000;
PR	14-SEP-2000;	2000US-02324001;	2000US-02324001;
PR	14-SEP-2000;	2000US-02330633;	2000US-02330633;
PR	14-SEP-2000;	2000US-02330634;	2000US-02330634;
PR	14-SEP-2000;	2000US-02330655;	2000US-02330655;
PR	21-SEP-2000;	2000US-02342223;	2000US-02342223;
PR	21-SEP-2000;	2000US-02342274;	2000US-02342274;
PR	25-SEP-2000;	2000US-02349977;	2000US-02349977;
PR	25-SEP-2000;	2000US-02349980;	2000US-02349980;
PR	26-SEP-2000;	2000US-02354884;	2000US-02354884;
PR	27-SEP-2000;	2000US-02356834;	2000US-02356834;
PR	27-SEP-2000;	2000US-02356836;	2000US-02356836;
PR	29-SEP-2000;	2000US-02363227;	2000US-02363227;
PR	29-SEP-2000;	2000US-02363677;	2000US-02363677;
PR	02-OCT-2000;	2000US-02370337;	2000US-02370337;
PR	02-OCT-2000;	2000US-02370338;	2000US-02370338;
PR	02-OCT-2000;	2000US-02370339;	2000US-02370339;
PR	02-OCT-2000;	2000US-02370440;	2000US-02370440;
PR	13-OCT-2000;	2000US-02399353;	2000US-02399353;
PR	13-OCT-2000;	2000US-02399377;	2000US-02399377;
PR	20-OCT-2000;	2000US-02409600;	2000US-02409600;
PR	20-OCT-2000;	2000US-02412121;	2000US-02412121;
PR	20-OCT-2000;	2000US-02417855;	2000US-02417855;
PR	20-OCT-2000;	2000US-02417856;	2000US-02417856;
PR	20-OCT-2000;	2000US-02417877;	2000US-02417877;
PR	20-OCT-2000;	2000US-02418008;	2000US-02418008;
PR	20-OCT-2000;	2000US-02418009;	2000US-02418009;
PR	20-OCT-2000;	2000US-02418266;	2000US-02418266;
PR	01-NOV-2000;	2000US-02446171;	2000US-02446171;
PR	08-NOV-2000;	2000US-02464755;	2000US-02464755;
PR	08-NOV-2000;	2000US-02464757;	2000US-02464757;
PR	08-NOV-2000;	2000US-02465225;	2000US-02465225;
PR	08-NOV-2000;	2000US-02465225;	2000US-02465225;
PR	08-NOV-2000;	2000US-02465226;	2000US-02465226;
PR	08-NOV-2000;	2000US-02465227;	2000US-02465227;
PR	08-NOV-2000;	2000US-02465228;	2000US-02465228;
PR	08-NOV-2000;	2000US-02465232;	2000US-02465232;
PR	08-NOV-2000;	2000US-02465238;	2000US-02465238;

PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	06-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases -

Claim 4; SEQ ID No 431; 1180pp; English.

The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. Influenza). The polynucleotides of the invention can also be used in gene therapy. AAS40795-AAS41684 represent cDNA sequences encoding for the novel human enzyme polypeptides of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct/sequences.

Sequence 629 BP; 189 A; 151 C; 124 G; 162 T; 3 other;

VS

```
Query Match      25.4%; Score 575.2; DB 22; Length 629;
Best Local Similarity 97.3%; Pred. No. 9.6e-155;
Matches 605; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

Qy 577 atcagggtatttagagatgtggttaccacgagctgagattaaacacctactcttgaca 636
Db 2 attcgccagaggagatgtggtttaccacgagctgagattaaacacctactcttgaca 61

Qy 637 agctcgctgcgaaggaggttaaacatggagaaactactatgtccagcctatttgcacacat 696
Db 62 agctcgctgcgaaggaggttaaacatggagaaactactatgtccagcctatttgcacacat 121

Qy 697 ccaggagtcagtttattactggaagtatcagatcacacacggacttcaacattctatca 756
Db 122 ccaggagtcagtttattactggaagtatcagatcacacacggacttcaacattctatca 181

Qy 757 taagacctaccacacccaaactgtttacacctggaacatgccaccctactcagaaactga 816
Db 182 taagacctaccacacccaaactgtttacacctggaacatgccaccctactcagaaactga 241

Qy 817 agggagttggatatcaacgcgatgtgctcgaaatggcacttgggtttttacagaaaaag 876
Db 242 agggagttggatatcaacgcgatgtgctcgaaatggcacttgggtttttacagaaaaag 301

Qy 877 aatgcattccacacagagagatttgatacctttttt-ggttccttttgggaagtggg 935
Db 302 aatgcattccacacagagagatttgataccttttttgggttccttttgggaagtggg 361

Qy 936 gattactatacacactacaaatgtgacagtcctggtggtggtgctgactgtatgaa 995
Db 362 gattactatacacactacaaatgtgacagtcctggtggtggtgctgactgtatgaa 421

Qy 996 aacgacaatgctgctgggactatgacaatggcgtatatactcacacagatgtacactca 1055
Db 422 aacgacaatgctgctgggactatgacaatggcgtatatactcacacagatgtacactca 481

Qy 1056 agagtacagcaaatcttagcttccataaccccaacaaagcctatatttttatattgcc 1115
Db 482 agagtacagcaaatcttagcttccataaccccaacaaagcctatatttttatattgcc 541

Qy 1116 tatcaagctgttattccaccactgcaagctcctggcaggtatttogaacatacagatcc 1175
Db 542 tatcaagctgttattccaccactgcaa-ctcctggcaggtatttccaacatacagatcc 600

Qy 1176 attatcaacataaacagagaa 1197
Db 601 attatccacntaacggangaa 622

RESULT 12
AAH99356
ID AAH99356 standard; cDNA; 562 BP.
XX
AC AAH99356;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:191.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnary; antiulcer; osteopathic; eczema;
KW dermatologicai; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
```

[illegible]

PR 08-NOV-2000; 2000US-02456613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-476224/51.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the respiratory system including respiratory
PT cancers and also for testing and detection e.g. diagnosis -
XX
XX Disclosure; SED ID No 1074; 546pp; English.
XX
XX The present invention relates to the isolation of novel human
CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful for preventing, treating and/or prognosing
CC disorders related to the respiratory system including throat
CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC the respiratory tissues e.g. lung cancer. The polynucleotide sequences
CC of the invention are useful in gene therapy and antisense therapy.
CC AAS28161-AAS28764 represent genomic sequences encoding for novel
CC human respiratory antigens.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1615 BP; 321 A; 458 C; 493 G; 343 T; 0 other;

Query Match 9.3%; Score 210.4; DB 22; Length 1615;
Best Local Similarity 59.3%; Pred. No. 1.2e-49;
Matches 377; Conservative 0; Mismatches 256; Indels 3; Gaps 1;

Qy 1384 gggctgtaggcttctgtcatagccactctgaaaaacaagggaacagtgtgtaaggac 1443
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ggggctaggcttctgtccacagtccctgtctcagcgaaagcaagcggaagccggggcac 60

Qy 1444 ctgtgcacatcactgactggtgtaccacactctcatttcaactggctgaaggacagatgatg 1503
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 tgatgcacatcactgactggtgtaccgagccctggtgggtctggtgaggtgtaccacctcag 120
Qy 1504 aggcattcaactagatggtctatgatctctgggagaccataagtgagggtctctcgtcac 1563
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 cagccgatgggttagtggtctacgctgtggtccgcatcagcagggcggtggtcctcac 180
Qy 1564 cccgagtagatatttgcataaaccattgacccatatacaccaggcaaaaaatggctcct 1623
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 cagcagcagagatcctgtgcacacattgacccactctacacccattgccagcatggctccc 240
Qy 1624 gggcagcaggtatgggtatgtggaacctgcaactccagtcagccatcagatgacgacct 1683
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 tggaggcggtcttggtcatcttggaacacgcgcgtgacgtgacctcccggtgggtgagt 300
Qy 1684 ggaattgcttacagaaaatcctgtgtacagcagctgggtgtccccctcagttctttcagca 1743
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 ggaagctgtgacagagagaccccggtatggcgattggatccacccgcagacactggcca 360
Qy 1744 acctgggaccgaaccggtggcacaatgaacgagatcacctcgtcaactggcaaaagtgtat 1803
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 ccttcccggtgtagctggtggaacctgggaacgaatggcc---agtgctcccgagccgtgt 417
Qy 1804 ggttttcaacatcacagccgaccatatagaggggtggacctatataacaggtatcccg 1863
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 ggtcttcaacatcagtgctgaccttatgaacggaggagacctggtgccagcgccgtg 477
Qy 1864 gaatcgtgaagaagctctctacaggtgtcactcagttcaacaaaaactgcagtcgcgttca 1923
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 478 atgtgtccgcacctgtggtcgtcgtggtggcggaataacacgcagacatcccggtac 537
Qy 1924 ggtatcccccaagacacccagaaagtaacctaggctcactgaagggtctagggacctat 1983
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 538 gctaccagctgagaacccccgggtcactcctgacttttaattgggtgtgcttggggccct 597
Qy 1984 ggtatagagagaaacaccaaagaaaagaaagcacaagca 2019
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 598 gggccagtgtaggaagagaggaggaggaagga 633
RESULT 14
ID AAA02370 standard; cDNA; 1082 BP.
XX
XX AAA02370;
XX
XX 19-MAY-2000 (first entry)
XX
XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2361.
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
XX 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX

PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI	Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI	Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI	Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX	WPI; 2000-126369/11.
DR	
XX	
PT	Polyucleotide library used to determine cancerous states of mammalian
PT	cells -
XX	
PS	Claim 1; Page 938; 1097pp; English.
XX	
CC	AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC	libraries constructed from human colon cancer cell lines. The present
CC	invention also describes a method of detecting differentially expressed
CC	genes correlated with a cancerous state of a mammalian cell, comprising
CC	detecting at least one differentially expressed gene product in a test
CC	sample derived from a cell suspected of being cancerous, where detection
CC	of the differentially expressed gene product is correlated with a
CC	cancerous state of the cell from which the test sample was derived.
CC	The polynucleotides sequences can be used in a method for detecting
CC	differentially expressed genes correlated with a cancerous state of a
CC	mammalian cell. The polynucleotides can also be used as probes for
CC	detecting and mapping related genes. They can be used in diagnosis and
CC	prognosis of diseases and disorders (e.g. identification of
CC	pre-metastatic or metastatic cancerous states, stages of cancer, or
CC	responsiveness of cancer to therapy). This is particularly for breast
CC	cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC	negative breast cancer, lung cancer, and colon cancer.
XX	
SQ	Sequence 1082 BP; 230 A; 182 C; 155 G; 190 T; 325 other:

	Query Match	6.2k; Score 139.6; DB 21; Length 1082;
	Best Local Similarity 93.0k; Pred. No. 2.1e-29;	
	Matches 187; Conservative 0; Mismatches 9; Indels 5; Gaps 4;	
QY	1493 acagattgatgagacattcaacta-gatggctatgatctcggagaccataagttagg 1551 	
Db	268 acagattgatgagacattcaactaggatggctatgatctcggagaccataagttagg 327 	
QY	1552 gtcttcgctcaccccgagttagatatatttgcataaca-ttgcaccatatacaccaaaggca 1610 	
Db	328 gtcttcgctcaccccgagttagatatatttgcatatgtacanttgaccatatacaccaaaggca 387 	
QY	1611 aaaaatggcttcctgggcagcaggctat-gggatctggaacactgcaatccagtcagccaat 1669 	
Db	388 aaaaatggcttcctgggcagcangctatggggatctggaacactgnaatccaatcagccaat 447 	
QY	1670 --cagagtgcagcactggaaa 1688 	
Db	448 tcaaagagggcagcactggaaa 458 	

RESULT	15
ABL12859	
ID	ABL12859 standard; cDNA; 1740 BP.
XX	
XX	
AC	ABL12859;
XX	
XX	
DT	26-MAR-2002 (first entry)
DT	
XX	
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 33059.

KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
XX	WO200171042-A2.
XX	
XX	27-SEP-2001.
XX	

PF	23-MAR-2001; 2001WO-US09231.	
XX		
XX	23-MAR-2000; 2000US-191637P.	
PR	11-JUL-2000; 2000US-0614150.	
XX		
XX	(PEKE) PE CORP NY.	
PA		
XX		
PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX		
DR	WPI; 2001-656860/75.	
DR	P-PSDB; ABB68756.	
XX		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	genes from Drosophila and for elucidating cell signalling and cell-cell	
PT	interactions -	
XX		
PS	Claim 1; SEQ ID NO 33059; 2lpp + Sequence Listing; English.	
XX		
CC	The invention relates to an isolated nucleic acid detection reagent	
CC	capable of detecting 1000 or more genes from Drosophila. The invention	
CC	useful in developmental biology and in elucidating cell signalling and	
CC	cell-cell interactions in higher eukaryotes for the development of	
CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA	
CC	sequences (ABL01840-ABL16175) and the encoded proteins	
CC	(ABB57737-ABB72072).	
CC	The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences .	
XX		
XX	Sequence 1740 BP; 469 A; 420 C; 441 G; 410 T; 0 other;	

Query Match	5.7%	Score 129;	DB 23;	Length 1740;
Best Local Similarity	48.5%;	Pred. No. 3.1e-26;		
Matches 502;	Conservative 0;	Mismatches 500;	Indels 33;	Gaps 4;
QY	547	agcccaatcattttcatcctagcggatgacacaggatttagagatgtgggtttaccacg	606	
Db	80	aaccaataatagtataactaatacgtacatggcagatgaacgatgcagttttcagt	139	
QY	607	gatic--tgagattaaacacctaactcttgacaagctcgctgcgaaggaggtataaac	663	
Db	140	gtcgaatcagatcctcacacgaataatagatgtttggcctacaatggtatcctgttga	199	
QY	664	agaactactgtccagccctatttgcacacatccaggagtcagtttattactggaagt	723	
Db	200	ataagcattatgttcgaacctatgcacacatccagagctactctgcactcggaagt	259	
QY	724	atcagatcacacccggacttcaacattctatcaataagacctcacccaacctgtttac	783	
Db	260	atccatacacacaggcatgcagcacttggattatccacgatgagccgtgggccccttc	319	
QY	784	ctctggacaatgcacccctaactcagaactgaagggttggttatccaacgcata	843	
Db	320	ctcagcagaaacgtcttatgccgaattttccggggtgagcggtactcgaccatttgg	379	
QY	844	tcggaaaatggcacttgggtttttacagaaaagaatgcatagccaccacagaagagtttg	903	
Db	380	tggcgaagtggcacttgggtcttggcgaggatctcacgcaccaaatgcgtggattcg	439	
QY	904	ataccttttggttcccttttgggaagtggggtattactatacacactacaaaatgtgaca	963	
Db	440	acatacactttggctactacaatggctacattgattactacgacctcaggtgcgaatgc	499	
QY	964	gtcctgggaatgtgtggctatgacttgcataaaacgacaatgctg---ctgggactatg	1020	
Db	500	tcgatagaactaactcggctggactgcgatttcccgcggtgatctggagccgtgcgccgaag	559	
QY	1021	acaatggcatatactcacacagatgtacactcagagagtacagcaaatcttagctccc	1080	
Db	560	cgaatggcacttacgcgcacgaagcccttacttctcggaagcaaaaggataattgagcaac	619	

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 17:15:04 ; Search time 69.6 seconds
(without alignments)
759.327 Million cell updates/sec

Title: US-09-495-823-7
Perfect score: 3012
Sequence: 1 MAPRCAGHPPPPSPQACVC.....VPRYPKPPSRNPLNGGV 550
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1424.5	47.3	533	1 KJHUAB	N-acetylgalactosam
2	1414.5	47.0	535	1 A44475	N-acetylgalactosam
3	1333	44.3	473	2 I54210	N-acetylgalactosam
4	472	15.7	521	2 H87394	sulfatase family p
5	454	15.1	497	2 AF0306	probable secreted
6	429	14.2	536	2 D83622	arylsulfatase PA01
7	422.5	14.0	522	1 KJHUG6	N-acetylgalactosam
8	416.5	13.8	567	2 A37362	arylsulfatase (EC
9	413	13.7	583	1 KJHUAC	steryl-sulfatase (
10	412.5	13.7	533	2 S69336	arylsulfatase (EC
11	407.5	13.5	589	2 I37187	arylsulfatase E (E
12	399	13.2	593	2 I37186	arylsulfatase D (E
13	398	13.2	551	2 C91220	arylsulfatase (imp
14	398	13.2	551	2 E86066	arylsulfatase (imp
15	395.5	13.1	537	2 F95993	probable arylsulfa
16	394	13.1	551	2 S30691	arylsulfatase (EC
17	393	13.0	555	2 D95377	probable cerebrosi
18	393	13.0	577	2 T45348	arylsulfatase (EC
19	378	12.5	507	1 KJHUA	cerebroside-sulfat
20	364	12.1	506	2 A54190	cerebroside-sulfat
21	357.5	11.9	551	2 S07089	arylsulfatase (EC
22	356	11.8	464	2 B35159	arylsulfatase (EC
23	349.5	11.6	535	2 AF0103	probable sulfatase
24	349.5	11.6	551	2 S01793	arylsulfatase (EC
25	344.5	11.4	303	2 T15835	hypothetical prote
26	336	11.2	545	2 F95325	probable arylsulfa
27	325	10.8	571	2 B85726	probable sulfatase
28	324	10.8	571	2 E64903	arylsulfatase homo
29	324	10.8	571	2 G90891	probable sulfatase

30	319.5	10.6	787	2	B70643	probable sulfatase
31	314.5	10.4	557	2	AF0370	probable sulfatase
32	272	9.0	495	2	AD3523	choline-sulfatase
33	257.5	8.5	517	2	AB0102	probable sulfatase
34	256	8.5	452	2	T29618	hypothetical prote
35	248.5	8.3	552	1	KJHUGU	N-acetylglucosamin
36	247	8.2	649	2	S43229	arylsulfatase (EC
37	243.5	8.1	970	2	E70533	probable sulfatase
38	243	8.1	497	2	C91206	probable sulfatase
39	243	8.1	497	2	E86052	probable sulfatase
40	241	8.0	497	2	G65169	probable sulfatase
41	221.5	7.4	503	2	E83642	choline-sulfatase
42	219.5	7.3	709	2	T16584	hypothetical prote
43	213.5	7.1	787	2	B70535	probable sulfatase
44	212	7.0	538	2	F83354	probable sulfatase
45	205.5	6.8	550	1	KJHUID	iduronate-2-sulfat

ALIGNMENTS

RESULT 1

KJHUAB

N-acetylgalactosamine-4-sulfatase (EC 3.1.6.12) precursor [validated] - human

N:Alternate names: arylsulfatase B (ASB); chondroitinase; chondroitinsulfatase; G4S;

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1993 #sequence_revision 27-Oct-1995 #text_change 08-Dec-2000

C:Accession: S35990; S33307; A35078; A45659; A42449; B42449; I54217; A56865

R:Peters, C.W.B.

Submitted to the EMBL Data Library, March 1993

A:Reference number: S35990

A:Accession: S35990

A:Molecule type: DNA

A:Residues: 1-533 <P>

A:Cross-references: EMBL:X72735; NID:g289009; PIDN:CAA51272.1; PID:g825628

R:Modaresi, S.; Rupp, K.; von Figura, K.; Peters, C.

Biol. Chem. Hoppe-Seyler 374, 327-335, 1993

A:Title: Cloning and expression of the human arylsulfatase B gene.

A:Refer

A:Access

A:Molecu

A:Resid

A:Cross

A>Note:

R:Peter

J. Biol

A:Title

A:Refer

A:Access

A:Molec

A:Resid

A:Cross

A>Note:

R:Litjel

Biochem

A:Title

A:Refer

A:Access

A:Molec

A:Resid

A:Cross

A>Note:

A:Note:

R:Jin,

Am. J.

A:Title: Mucopolysaccharidosis type VI: identification of three mutations in the aryl

city.

A:Reference number: A42449; MUID:92197625

A:Accession: A42449

A:Molecule type: mRNA

EMBL:X72738; EMBL:X72739;

Isidorf, M.; Vingron, M.; M

cloning and expression of

: PID:g179077

equencing

Hopwood, J.J.

maturation and isolation o

: PID:g236698

NCBIP:57778)

the mature protein, were d

ewhere between residue 450

102 C5) >45%

Db 39 -----AGAG-----ADRPPLHVFVLADLGNWDSVPHGSNIRTPHLDLAAGGVLLD 85

Qy 115 NYVQPICTSRSFQITGKYQIHTGLQHSIRTPQPNCLPLDNATLPQKLKEVGYSTHMV 174

Db 86 NYVQPICTSRSFQITGKYQIHTGLQHSIRTPQPNCLPLDNATLPQKLKEVGYSTHMV 145

Qy 175 GKWLHGYRKECMPTRRGFDTFGSLGSDYTHYKC---DSPGM--CGYDLYENDNAA 229

Db 146 GKWLHGYRKECMPTRRGFDTFGSLGSDYTHYKC---DSPGM--CGYDLYENDNAA 205

Qy 230 WYDNGIYSTQMTYTORVQOITLASHNPKPIFLAYQAVHSPLOAPGRYFEHYSIININ 289

Db 206 TGYKN-MYSNIFTERATLITSHPPKPLFLYALQSHVHEPQVPEYLPKDYFIDOKN 264

Qy 290 RRYAAMLSCLDEAINNVTLALKTYGYNNSSIIYSDNGQPTAGSNMPLRSGKGTW 349

Db 265 RHYVAGVMSLMDAENVNTAALKSHGLWNNTVFIFSDNGQTLAGGNMPLRGRKWSLW 324

Qy 350 EGGIRAVGFVHSPLLKNGKIVCEPVHITDWTPTLISLAEGQDDEDIQLDGYDIWETISE 409

Db 325 EGGIRAVGFVHSPLLKNGKIVCEPVHITDWTPTLISLAEGQDDEDIQLDGYDIWETISE 384

Qy 410 GLRSPRYDILHNDIPY-----TRAKNGSWAAGYIWNNTAIOAIRVQHWKLLT 458

Db 385 GSPSPRELHNDIPNVDISPCPKSLAPAKDDSSHPAYLAFWISLHAIRHGNWKLTT 444

Qy 459 GNGYSOWPVPQSFNSLGNRWHNERITSS--TGKSYWLFNITADPYRVDLSNRYPGIV 516

Db 445 GYPCGCGWFPP-----PSYNDISAIPSSDPPTKTLWPFIDQDPEERHDLSDRYPHIV 497

Qy 517 KLLRLRSQNKNTAVPVRYPPKDPSPRNPLNG 548

Db 498 EQLLSRLQFYHKKHVSVPVHFAQDPDPCDPRCKGTG 529

RESULT 3

I54210

N-acetylglucosamine-4-sulfatase (EC 3.1.6.12) precursor - rat (fragment)

N:Alternate names: arylsulfatase (EC 3.1.6.1) B [misidentification]

C:Species: Rattus norvegicus (Norway rat)

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 02-Jun-2000

C:Accession: I54210

R:Kunieda, T.

Genomics 29, 582-587, 1995

A:Title: Mucopolysaccharidosis type VI in rats: Isolation of cDNAs encoding arylsulfatases

A:Reference number: I54210; MUID:96121368

A:Accession: I54210

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-473 <RES>

A:Cross-references: GB:D49434; NID:g1065603; PIDN:BAA08412.1; PID:g1089794

C:Genetics:

A:Gene: ARSB

C:Superfamily: animal sulfatase

C:Keywords: sulfuric ester hydrolase

F:31/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 44.3%; Score 1333; DB 2; Length 473;

Best Local Similarity 53.9%; Pred. No. 2.6e-94;

Matches 255; Conservative 70; Mismatches 122; Indels 26; Gaps 5;

Qy 94 HGSIKTPTLDKLAAGVKKLENYVQPICTPSRSQFITGKYQIHTGLQHSIRTPQPNCL 153

Db 3 HGSVIRPHLDALAAGVLDNYVQPICTPSRSQFITGKYQIHTGLQHSIRTPQPNCL 62

Qy 154 PLDNATLPQKLKEVGYSTHMVGVKWHLGFYRKECMPTRRGFDTFGSLGSDYTHYKC--- 209

Db 63 PLDEKLLPQLKDGSSSTMVGVKWHLGFYRKECMPTRRGFDTFGSLGSDYTHYKC--- 122

Qy 210 -YKCDSPGCMGYDLYENDNAAWDYDNGIYSTQMTYTORVQOILASHNPKPIFLYIAQAV 268

Db 123 PIECLNTRCALDRDGEPEAKETD-IYSTNIETKRAITLIANHPPEKPLFLYLAQSV 181

Qy 269 HSPLQAPGRYFEHYSIININRRRYAAMLSCLDEAINNVTLALKTYGYNNSSIIYSDN 328

Db 182 HDPLQVPEEYMEPYDFIQDKHRRIRYAGVMSILLDEAVGNVTALKSRGLWNNTLIFSTDN 241

Qy 329 GGQPTAGGSNWPPLRSGKGTWEGGIRAVGVHSPVPLKNGKIVCEPVHITDWTPTLISLA 388

Db 242 GGQTRSGNWNPLRGRKGTWEGGIRAVGVHSPVPLKNGKIVCEPVHITDWTPTLISLA 301

Qy 389 EGQIDEDIQDGDYDIWETISEGLRSRPRVDILHNDIPY-----TKAKNGSWAAG 437

Db 302 GGSTHGTKPLDGDVWETISEGSPSPRVVELLNLNDIPDFDGLPCPGKNTYTPKNDSPLE 361

Qy 438 YGIWNTAIOAIRVQHWKLLTGNPGYSDWVPPQSFNSL--PNRWHNERITSSYTKSVWL 495

Db 362 HSAFNTSIHAGIRYKKNWKLTYGPGCGYWFPPPSQSNISEVPS-----VDSPTKTLWL 414

Qy 496 FNITADPYRVDLSNRYPGIVKLLRLRSQNKNTAVPVRYPPKDPSPRNPLNG 548

Db 415 FDIRDPEERHDSREPHIVQNLISRLQYHHSVSPYFPLDPRCDPRCKGTG 467

RESULT 4

H87394

sulfatase family protein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: H87394

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: H87394

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-521 <STO>

A:Cross-references: GB:AE0055673; NID:g13422494; PIDN:AAK23156.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC1172

Query Match 15.7%; Score 472; DB 2; Length 521;

Best Local Similarity 26.6%; Pred. No. 3.2e-38;

Matches 142; Conservative 69; Mismatches 193; Indels 130; Gaps 16;

Qy 68 PSTTSTSQPHLIFILADDDQGFDRV-----GYHSEIKTPTLDKLAAGVKKLEN-YYVQPI 121

Db 25 PPADGQRPNNIVFTLADDLGYNDITLNGGVAGGVTPPTPAIDSTAKEGVTFANGYSGNAT 84

Qy 122 CTPSRQFITGKYQIHTGLQHS-----IIRPTQPN----- 151

Db 85 CAPSRAIMTGRYATRFGEFTPTPVAFSRVVGHGADPLHPSRFNAEYKMPKDNVL 144

Qy 152 CLPLDNATLPQKLKEVGYSTHMVGVKWHLGFYRKECMPTRRGFDTFGSLGSDYTHYKC---GD 205

Db 145 AVPAAEVTIAEALTKGYHTHLGKWHLGGV-KGSRPEDQGFDESGLGFMAGAAALFAPVGD 203

Qy 206 YYTHYKCDSPGCMGYDLYENDNAAWD-----YDNGIYSTQMTY 244

Db 204 -----PGV-----ESRQDWDPIDKFLWGAAPFAVQFNGGKLFNFSHYMTDYLTD 248

Qy 245 RVQOILASHNPKPIFLYIAQAVHSPLOAPGRYFEHYSIININRRRYAAMLSCLDEAI 304

Db 249 EAVRAIDA-KNRPFEMFLAYNAVHTLQAPKADYDALSHIKDRMRYAAMVNRNLDNRV 307

Qy 305 NNVTALKTYGYNNSSIIYSDNGQPTAG--GSMNPLRSGKGTWEGGIRAVGVHSP 362

Db 308 GKVLQALKRLDNDTLVIFTSNGGANYICLPDINKPYRGWKATFEKGKIVPFLLRWP 367

Qy 363 LLKKNKGVCKECPVHITDWTPTLISLAEGQDDEDIQLDGYDIWETISEGLRSRPRVDILHNI 422

Db 368 EQLPAGAVYRSPGVHVDIFATAAGAAGAPTTPKDRVLDCVDL-----VFFVKQSSG 418
Qy 423 DPIYTKAKGNSWAAGYGIWNTAIOAISAIRVQHWKLLTGNPGYSDWVPPQSFNSLGNRWHN 482
Db 419 DP--HKAI-----FWRSGGYKTVLAGWKQVAK-----PNK---- 449
Qy 483 ERITSSGKSWLNFNITADPYRVDLSNRYPGVIGVKLLRLRLSQFNKTAVPVRY 536
Db 450 -----TWLFDLSDTPFERQELSKARPEKLEKREMAQALDQGMKWTWP 493

RESULT 5
AF0506
A:Title: Complete genome sequence of *Salmonella enterica* serovar Typhi
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A:Note: This species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AF0506
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; Farrar, M.; S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi
A:Reference number: AB0502; PMID:11677608
A:Accession: AF0506
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-497 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01190.1; PID:gl6501319; GSPDB:GN00176
C:Genetics:
A:Gene: STY0042

Query Match 15.18; Score 454; DB 2; Length 497;
Best Local Similarity 26.38; Pred. No. 7.1e-27;
Matches 138; Conservative 77; Mismatches 189; Indels 120; Gaps 16;

Qy 62 AGEKLEPSTTTS-QPHLIFILADDOGRDVGHYGSEI-KPTLDKLAAGVKLENYV- 118
Db 16 AGCVLSAQAATKRNPLVILADDLGVLATYGHVKTPTNIDKLAQEGVKETDIYAP 75
Qy 119 QPCTPSSQITGKYQIHTGLQHSIIRTPQPNCLPDNATLPQKLEKVGYSTHWGKWH 178
Db 76 APLCSPSRAGLLTGRMPRTGIR-SWIPEGKDVAGLGRNELTIANLLKQOQYDTAMMGKLI 134
Qy 179 L--GFYRKECMPTRRGDTFFGSLGSDYYTHYKCDSPGCMGVDLYENDNA----- 228
Db 135 LNAAGDRTPQAKDMGFDY---TLVNPAGVTDATL-----DNAKERPRYG 178
Qy 229 -----AWDYDNGIY-----STQMYTORVOQILASHNPTKPIFLYAYQAVHSPLOAPG 276
Db 179 VVHPGTW-IRNGHIGRADKMSGEFVSSEVYNLDNKKDDNPFFLYVAFTEVHSPLASPK 237
Qy 277 RYFHYRSIIINRR-----YAMLSCLEAINNVTLAKTYGYF 317
Db 238 KYLDMYSQMYTDYQKHQPDLYFGDWADKPWRGTGEITYANISYMDYEQVGKVLDKIKAMGEE 297
Qy 318 NSIIIISSDNGG-----OPTAGSNWPLRSGKTYWEGGIRAGVFV-----HSPLL 364
Db 298 DNTVIFSDNGPVYTRARKYVELKLGETDGLGRKDNLWEGGIRVPAITKYGKHIP-- 355
Qy 365 KNGKTYCKEPVHITDWPYTLISLAEGQIDEDIQLDGVDIWEITISEGLRSPRVDILHNIDP 424
Db 356 --QGMVTDTPVYGLDWLPTLANMMDFKLPTDRTYDQGSLVPLDKTLKQKPLIFGIDM 413
Qy 425 IYTKAKGNSWAAGYGIWNTAIOAISAIRVQHWKLLTGNPGYSDWVPPQSFNSLGNRWHNER 484
Db 414 PFQDDPTDEW-----AIRDGDWKN----- 432
Qy 485 ITSSTGKSWLNFNITADPYRVDLSNRYPGVIGVKLLRLRLSQFNK 528
Db 433 IIDRQNKPKYDNLKTRDFTLNQIGKQPKQLEKQYKFLAYK 476

RESULT 6
D83622
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83622
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337
A:Accession: D83622
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-536 <STO>
A:Cross-references: GB:AE004456; GB:AE004091; NID:g9946013; PIDN:AAG03573.1; GSPDB:GN00176
A:Experimental source: strain PA01
C:Genetics:
A:Gene: atsA; PA0183
C:Superfamily: animal sulfatase

Query Match 14.28; Score 429; DB 2; Length 536;
Best Local Similarity 24.98; Pred. No. 6.5e-25;
Matches 144; Conservative 71; Mismatches 130; Indels 174; Gaps 20;

Qy 73 TSQPHLIFILADDOGRDVGHYGSEI-KPTLDKLAAGVKLENYVQIPTPSRSQFTIG 132
Db 2 SKRPNFLVIVADDLGFSDIGAFGEIATPNDALAIAGLITDFTASTCSPTRSMLL-- 59
Qy 133 KYQIHTGLQHSII-----RTPQPNCLPDNATLPQKLEKVGYSTHWGKWH 177
Db 60 -----TGTGDHIAIGTMAEALTPLEGKPGYEGHNLNRVVALPELLREAGYQTLMAKWH 114
Qy 178 HLGFYRKECMPTRRGDTFFGSLGSDYYTHYKCDSPGCMGYD-----LYEN 225
Db 115 HLGL-KPEQTPHARGFERFSLLFGAAN--HYGFEPP-----YDESTPRILKGTALYVE 166
Qy 226 DNAAMD-YDNGIYSTQMYTORVOQILASHNPTKPIFLYAYQAVHSPLOAPGRYFEHYRS 284
Db 167 DERYLDTLPEGYSSDAFGDKLQYLKERDQSRPFPAVLPSAPHWLPQAPREIVEKYRG 226
Qy 285 -----IININR-----RRYAAAM 296
Db 227 RYDAGPEALROERLARLKEGLVEADVEAHPVLAALTREWEALEDEERAKSARAMEVYAA 286
Qy 297 LSCLEAINNVTLAKTYGFYNNIIIISSDNGG-----PTAG----- 335
Db 287 VERMDWNGIRVVDYLRROGELDNFTVLFMSDNGAEGALLEAFKPGPDLLGLDRHYDNS 346
Qy 336 -----GSNW-----PLRSGKTYWEGGIRAGVFHSPVLSPLKNGTKVCKEVPVH 376
Db 347 LENIGRANSYVYVYVPRWAQAATAPSRLYKAPTQGGIRVPALVRYPRLSRQCAISHAFAT 406
Qy 377 ITDWPYTLISLAEGQIDEDIQLDGVDIWEITISE---GLRSPRVDILHNIDPIYTKAKNGS 433
Db 407 VMDVPTLLDLA-GVRHPGKWRGREAEPGRSWLGSWLSGTEAAHDENTV----- 457
Qy 434 WAAGYGIWNTAIOAISAIRVQHWKLLTGNPGYSDWVPPQSFNSLGNRWHNERITSSTGKSV 493
Db 458 --TG---WELFGMAIRQGDWKNV-----YLP----APVGATWQ----- 488
Qy 494 WLFNITADPYRVDLSNRYPGVIGVKLLRLRLSQFNKTA 531
Db 489 -LYDLARDPGEIHLADSQPGKLAELIEHWRKYVSETGV 526

RESULT 7
KJHUG6
N-acetylglactosamine-6-sulfatase (EC 3.1.6.4) precursor [validated] - human

Qy	411	LRSPRVDLINIDPIYTKAKNGSWAACGCTWNTAIQSRAIRVQHKKLLTGTNGPGVSDWVPPQ	470
Db	376	RLMDRPITYYRGDTLMA-ATLGQHKAHFTWTNS-----WENPQG---IDFCPGQ	422
Qy	471	SFSLGNPNRWHNERITSSSTGKSVMLFNITADPYERVDL---SNRYPGIVAKKLLRRLSQFN	527
Db	423	NVSCV---TTHN---LEDHKTLPFIHLGRDPCRFPFLSPASAEYQEALSRITSVVOHQ	476
Qy	528	KTAVPVRYPPKDRPSRPNLN	547
Db	477	EALVP-----AQQLN	487
RESULT 8			
A37362			
arylsulfatase (EC 3.1.6.1) precursor - sea urchin (Strongylocentrotus purpuratus)			
C:Species: Strongylocentrotus purpuratus (purple urchin)			
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Nov-1999			
C:Accession: A37362			
R:Yang, Q.; Angerer, L.M.; Angerer, R.C.			
Dev. Biol. 135, 53-65, 1989			
A:Title: Structure and tissue-specific developmental expression of a sea urchin aryls			
A:Reference number: A37362; MUID:89357267			
A:Accession: A37362			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-567 <YAN>			
A:Cross-references: GB:M28404; GB:M25815; NID:g161440; PID:g161441			
C:Superfamily: animal sulfatase			
C:Keywords: sulfuric ester hydrolase			
F:115/Modified site: 3-oxoalanine (Cys) #status predicted			
Query Match 13.8%; Score 416.5; DB 2; Length 567;			
Best Local Similarity 29.7%; Pred. No. 6.4e-24;			
Matches 120; Conservative 66; Mismatches 175; Indels 43; Gaps 14;			
Qy	58	LLAQAGEKLEPSTSTSQPHLIFLADDOGPRDVGYHGSEIKTP-TLDKLAAGVKL-EN	115
Db	53	LLGQTGQ----HRTAMTKPNVILLADDMGVGDLISVYGHPTQEPGFIDQMANQGLRFTQG	108
Qy	116	YVVOPICTPSRSQITKGVIQHTGL--OHSIIRTPQNCPLDNLTPQKLKEVGYSTHM	173
Db	109	YSGDSVCTPSRAIVTGRQPIRTGVYGEERFLPWTITGLPLYEVTTAEAMKAGCYTTGM	168
Qy	174	VGKWHLGYRKEC-----MPTRRGDFTFGSLGSGDYTHYKCDSPGMCYDILYENDNA	228
Db	169	VGKWHLGINENSSDGAHLPANRGFD-FVGHNLPFNG---SWRCDDTGL--HQDFPDNTA	222
Qy	229	AWDYDNGIYSTQMYTOR-VQOIL-----ASHNPTKPIFLYIAYQAVHSPLOAPGRFFE	280
Db	223	CFLYNSTSAQPFQHKLTOLLRDDTVGFIEDNVNPFMYVSFAHMTSL-----FS	276
Qy	281	HYRSIINIRRRYAAMLSCDLAEANNVTALKTYGFYNNISIIYSSDNG-----GOPTA	334
Db	277	SDDFCSCTSRRGYGNLREMDQAEIQIVTILVDNIDDTNTVIFTSDHGPHREYCGE---	333
Qy	335	GGSNMPLRGSKGTWYEGGIRAVGFVHSPLLKKNKTCVCKEPVHITDWPYTLISLAEGIDE	394
Db	334	GGDANVRFGKGQSWEGGHRIPYIVYVPGTISPG-VSHEIVTSMIIATAVNLGGSQLP	392
Qy	395	DIQLDGDYDIWTISIEGLSPRVDILHIDPIYTKAKNGSWAAGY	438
Db	393	DRIVDGGCKLSVLEGASPHDDFFYYCKDTLMAVRVGKYKAHF	436
RESULT 9			
KJHUNC			
steryl-sulfatase (EC 3.1.6.2) precursor - human			
N:Alternate names: arylsulfatase C; steroid sulfatase (STS); steryl-sulfate sulfohyd			
C:Species: Homo sapiens (man)			
C:Date: 21-Mar-1990 #sequence_revision 27-Oct-1995 #text_change 11-Mar-2000			

Db	128	TSAYSQSSSPTRATILITGOYSIH-----HGILMPMPYQPGGLQ-GLTTLPLQLHDQYV	182
Qy	171	THMVGKWLHGYRKECMPTRRGDTF--FGSLLSGDYVTHYKCDSPGCMGYDLYENDNA	228
Db	183	TOAIGKWHMG-ENKESQPNQVDFDFGFSV---SDMYTEWR-----DVHVNPEV	229
Qy	229	AWDYNIGYSTQMYTOR-----VQOILASHNP-----	255
Db	230	ALSPDRSEYIKQLPFSKDDVHAVRGGQQAIAIDITPKYMEDLDQRMWMEYGVKFLDKWAKS	289
Qy	256	TKPIFLYIAQVHSPLOAGRFEHYRSI-----ININRRRYAAMLSCLDEAINNVTLAL	311
Db	290	DKPEFLYIGTRGCH-----FDNPNKAYAGSSPARTSYGDCVMVEMNDVFANLYKAL	340
Qy	312	KTYGFYNNIISSDNGGO---PTAGGSNWPLRGSGTYWEGGIRAVGFVHSPLLANKG	368
Db	341	EKNQDLNTLIVTSDNGPAEVPHPHRT--PFRGAGSGTWEGGVVPTFYW-----KG	393
Qy	369	TVCKEP-----VHITDWPYPLISAEQIDEDIQLDGYDIWETISEGLRSRPRVDILHNI	422
Db	394	MI--QPRKSGDGVDLADLFPTALDLA-GH-----PGAKVANLV-----PKTTFIDGV	437
Qy	423	DP-----IYTRAKGWSWAAGYGINWTAQSAIRVOHWKLLTGNPGYSDWV-PPQSFSLGP	477
Db	438	DOTSFFLGINGQSNRKAHEHYFL--NGLAAVRMDDEF-----YHVLIOQPYAYTOSGY	488
Qy	478	NRWNERITSGKSVWLFNITADPYERVDLSNRYPGVKKLLRRLSQFNKTAVPVRYPP	537
Db	489	OGGFTGVMTAGSSV--FNLYTDPQSDSIGVRHPGVPGLQTEMHAYWE--ILKAYPP	544
Qy	538	K 538	
Db	545	R 545	

RESULT 15
F95993
probable arylsulfatase (EC 3.1.6.1) [imported] - Sinorhizobium meliloti (strain 1021) me
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 14-Sep-2001
C;Accession: F95993
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: F95993
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-537 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC49614.1; PID:g15141101; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymb
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: aslA1; SMB20915
A;Genome: plasmid
C;Keywords: sulfuric ester hydrolase

Qy	65	KLEPSTSTST--QPHLIFILADDOG-FRDVGYHGSEI--KTPPTLDKLAEGVKLENYVQ	119
Db	43	KAQAASTGGSGRKPNIILMIMGDDIGWFNPSIYHRGMMGYRTPNIDRIGNEGAMFTDWYGE	102

Qy	120	PICTPSRSQFITKYQIHTGLQHSIIIRPTOPNCLPLDNATLPKRLKEVGYSTHMVGKWHL	179
Db	103	QSCTAGAAAITGOSPRTGLT-KVGLPGADIGIOPEDAIVAEALLKSLGATGQFGKNHL	161
Qy	180	GFYRKECMPTRRGDTFFGSLGSGDYVTHYKCDSPGCMGYDLYENDNAWDYDNGIYST	239
Db	162	G-DKDEFLPTAHGFDEFNGLY-----HLNAEE-----EPENPDYQDDPA-FRK	203
Qy	240	QMYTORVQOILASH--NPTKPI-----FLYTAYQAVHSPLOAGRFEHYRSI-	285
Db	204	RFGRGVIKATADGKIEDTGPLSVKRMETVDQEFLEALDFIDRKAKGDAPWFCYFNSTR	263
Qy	286	--ININRRRYAAMLSCL-----DEAINNVTLAKTYGFYNNIIIIYSSSNGGOPT	333
Db	264	MHVNTHLKPDSAGKTGLCVYPDGVNVEHDGHVGLKLLKLDLGLTENTIVVYTSNGAEVM	323
Qy	334	A--GGSNWPLRGSGTYWEGGIRAVGFVHSPLLKKNKGTVCKEPVHITDWPYT-----	383
Db	324	TWPDGGNTPFERKATNWEGGFRVPMCIWRPGVIKPGTIHNEPESHYDLIPTFCAAAAGEP	383
Qy	384	-----LISLAEGQIDEDIQLDGYDIWETISEGLR-SPRVDILHNIIDPIYTKAKNGSWAA	436
Db	384	DIVAKCLTGYAAGAKTFKVLHDGYNLMPFLSGSSNDAPRRDFLY-----	427
Qy	437	GYGIWNTAIQ--SAIRVOHWKLL---TGNPGYSDWVPQSFNSLGNPNRWNERITSTGKS	492
Db	428	-----WDDGELVAVRVQNNKVVFKSQDBEGIGVWRQP--FTEL-----RA	466
Qy	493	VWLFNITADPYERVDLSNRY-----PGIVKKLLRRLSQFNKTAVPVRYPP	537
Db	467	PLLFNLRSDPFERGDTSMYEYKFPFDRSFVVVPSQAVVAKWLESFKEF-----PIROKP	520

Search completed: July 23, 2002, 19:04:13
Job time: 6549 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:04:54 ; Search time 4089.32 Seconds
(without alignments)
11595.946 Million cell updates/sec

Title: US-09-495-823-8
Perfect score: 2266
Sequence: 1 cagcgtccgcacacgcgtc.....tgccaactgtgtgcgaattc 2266

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_to.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_to.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES
Result No. Score Match Length DB ID Description

1	2234.2	98.6	2253	6	AX206967	AX206967
2	1781.6	78.6	1800	6	AX206973	AX206973
c 3	1515	66.9	152861	9	AC104779	AC104779 Homo sapi
c 4	1515	66.9	165147	2	AC021342	AC021342 Homo sapi
c 5	1134.2	50.1	232951	2	AC091322	AC091322 Mus muscu
6	789.8	34.9	1871	9	AK027201	AK027201 Homo sapi
c 7	690.6	30.5	157043	2	AC013692	AC013692 Homo sapi
c 8	690.6	30.5	165147	2	AC021342	AC021342 Homo sapi
c 9	596.2	30.5	175645	2	AC105414	AC105414 Homo sapi
10	596.2	26.3	2940	6	AX206965	AX206965 Sequence
11	594.4	26.2	1710	6	AX206972	AX206972 Sequence
12	496.8	21.9	181923	2	AC011372	AC011372 Homo sapi
13	398.2	17.6	2805	3	AF109924	AF109924 Helix pom
14	379	16.7	1939	4	S48472	S48472 arylsulfa
15	368.2	16.2	2802	9	HUMARSBX	M32373 Human aryls
16	366.6	16.2	2228	9	HUMASFB	J05225 Human aryls
17	332.6	14.7	1603	10	RATARSB	D49434 Rat ARSB mR
18	303.8	13.4	122992	2	AC098604	AC098604 Rattus no
19	278.6	12.3	176341	2	AC105383	AC105383 Homo sapi
20	242.8	10.7	758	10	MMASLS	X92096 M.musculus
21	189.2	8.3	853	3	AF109925	AF109925 Helix pom
22	149.8	6.6	1428	10	BC020108	BC020108 Mus muscu
23	141.4	6.2	1777	3	AY071072	AY071072 Drosophil
c 24	138.4	6.1	16052	2	AC109105	AC109105 Rattus no
25	108.2	4.8	329	9	HSARYB5	X72739 H.sapiens g
26	106.6	4.7	114624	9	AC099485	AC099485 Homo sapi
27	106.6	4.7	145718	2	AC091369	AC091369 Rattus no
c 28	106.6	4.7	176356	2	AC013689	AC013689 Homo sapi
c 29	102.4	4.5	164478	2	AC008787	AC008787 Homo sapi
30	95.8	4.2	271	9	HSARYB2	X72736 H.sapiens g
31	95.8	4.2	94161	9	AC020937	AC020937 Homo sapi
32	95.8	4.2	119582	2	AC022110	AC022110 Homo sapi
33	95.8	4.2	164478	2	AC008787	AC008787 Homo sapi
34	76.2	3.4	1926	9	HSARYB1	X72735 H.sapiens g
35	76.2	3.4	187592	2	AC069033	AC069033 Homo sapi
36	75	3.3	766	9	S57777	S57777 N-acetylgl
37	67.8	3.0	293	9	HSARYB4	X72738 H.sapiens g
c 38	61.2	2.7	119582	2	AC022110	AC022110 Homo sapi
c 39	58	2.6	7218	6	I66494	I66494 Sequence 14
40	56	2.5	203291	2	AL596447	AL596447 Mus muscu
41	55.6	2.5	32037	2	AC018283	AC018283 Drosophil
c 42	55.6	2.5	179649	3	AC010045	AC010045 Drosophil
43	55.6	2.5	290827	3	AE003522	AE003522 Drosophil
c 44	55.4	2.4	87916	2	AC014974	AC014974 Drosophil
45	55.4	2.4	158520	3	AC016021	AC016021 Drosophil

ALIGNMENTS

RESULT	1					
AX206967	AX206967	Sequence 8 from Patent WO0155411.	2253 bp	DNA	linear	PAT 30-AUG-2001
LOCUS	AX206967	Sequence 8 from Patent WO0155411.				
DEFINITION	AX206967	Sequence 8 from Patent WO0155411.				
ACCESSION	AX206967	Sequence 8 from Patent WO0155411.				
VERSION	AX206967.1	GI:15394724				
KEYWORDS	.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 2253)				
AUTHORS	Glucksmann,M.A., Williamson,M., Rudolph-Owen,L.A. and Tsai,F.Y.					
TITLE	Human sulfatases					
JOURNAL	Patent: WO 0155411-A 8 02-AUG-2001;					
FEATURES	Millennium Pharmaceuticals, Inc. (US)					
source	Location/Qualifiers					
	1..2253					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	324..2123					
	/note="unnamed protein product"					
	/codon_start=1					
CDS						

Db 1861 CAGGAATCGTGAAGAAGCTCCTACGGAGCTCTCACAGTTCAACAAAACTGCAGTGC GG 1920
Qy 1921 tcaggtatcccccaagaccaccccaagtaaaccttagctcaatgagggctctaggac 1980
Db 1921 TCAGGTATCCCCCANNAGCCCCAGAGTAACCTTAGCTCAATGAGGGGTCTGGGGAC 1980
Qy 1981 catggtatagagagaaaccagaagaaagcaagcaaaatcaggctgagaaaaagc 2040
Db 1981 CATGTTATAAGAGAGAAACCAAGAAAAGAACCAAGCAAAATCAGGCTCAGAAAAAGC 2040
Qy 2041 aaagaaagc-aaaaaagaagaagaaacagcagagaagcaggtctcaggttcaactggc 2099
Db 2041 AAAAGAAAGCAAAAAAAGAAAGAAACACAGCAGAAAGCAGTCTCAGGTTCAACTGCC 2100
Qy 2100 atcagggttactgtgtgataagcacaaatatttctgttctgttgaacttttaactcagt 2159
Db 2101 ATTCAAGTGTACTGTGGATAGCACAAATATTTCTGTGGTTAAACATTTAATCAGT 2160
Qy 2160 tcttatcttctcatctgtttcttagtaaacccagcaaaatttgctcgataaatcgtctggc 2219
Db 2161 TCTATCTTTTCATCTGTTTCTTAGTAAACCCAGCAAAATTTGCTCGATAATATCGCTGC 2220
Qy 2220 ctaagcgtcaggtctgttttctcatcgtctgcccac 2252
Db 2221 CTAAGCGTCAGGCTGTGTTTTCATGCTGTGCCAC 2253

RESULT 2
AX206973 AX206973 1800 bp DNA linear PAT 30-AUG-2001
LOCUS Sequence 14 from Patent WO0155411.
DEFINITION AX206973
ACCESSION AX206973
VERSION AX206973.1 GI:15394729
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1800)
AUTHORS Glucksmann,M.A., Williamson,M., Rudolph-Owen,L.A. and Tsai,F.Y.
TITLE Human sulfatases
JOURNAL Patent: WO 0155411-A 14 02-AUG-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source 1..1800
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 534 a 434 c 423 g 409 t
ORIGIN

Query Match 78.6%; Score 1781.6; DB 6; Length 1800;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1795; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 324 atggctcccagggtgtgtcggggcatccgctcgccttctccacagggctgtgtctgt 383
Db 1 ATGGCTCCACAGGGGTGTGCGGGGATCCGCCCTTCGCCCTTCCACAGGCGCTGTCTGT 60
Qy 384 cctgaaagatgctagcaaatggggcgcgtgcgaggaattcgtgactcgtcctcact 443
Db 61 COTGAAAGATGCTAGCAATGGGGCGGTGCGAGGATTCGTGATCCTCTGCTCCTCACT 120
Qy 444 tatgttacctgtctctgggcagccttagaagaggaggaagggccttactagct 503
Db 121 TATGTTTACCTGTCTCTGGGCGAGGCCCTTAGAAGAGGAGGAAGAGGGGCCCTTACTAGCT 180
Qy 504 caagctgagagaaactagagccagcacaaacttccactccctccagcccatctcatttc 563
Db 181 CAAGCTGGAGAGAACTAGAGCCCGACACACTTCCACCTCCCGCCCATCTCATTTTC 240
Qy 564 atcctagcggatgatcaggggatttagagatgtggtttaccacgagatctgagattaaaaa 623

Db 241 ATCTAGCGGATGATCAGGGATTTAGAGATGTGGGTTACCACGGATCTGAGATTAAACA 300
Qy 624 cctactcttaagaagctcgcgtgcgaagagacttaaacctggagaactactatgtccagcct 683
Db 301 CCTACTCTTGACAAGCTCGCTGCGGAGGAGTAAACTTGGAGAACTACTATGTTCAGGCT 360
Qy 684 attgcacaccatccaggagtcagtttatctactggaagtatcagatcacacacggactt 743
Db 361 ATTTGCACACCATCCAGGAGTCAGTTTATTACTGGAAGTATCAGATACACACGGACTT 420
Qy 744 caacattctatcataagacctaccacaccccaactgtttacctctggacaaatgcaccccta 803
Db 421 CAACATTTCTATCAATAAGACCTACCCACCAACCTGTTTACCTCTGTGGAATGCCACCTA 480
Qy 804 cctcagaaactgaaggaggttgatatattcaacgcataatgctcgaaaaatgcacttgggt 863
Db 481 CCTCAGAAACTGAAGGAGGTGGATATTCAACGCATATGGTCGGAATGGCATTGGGT 540
Qy 864 ttltacagaaagaatgcatgcccaccagaagagatttgataaccttttttgggtccctt 923
Db 541 TTTTACAGAAAGAAATGCATGCCACACAGAGAGGATTTGTATACCTTTTGGTTCCCTT 600
Qy 924 ttgggaagtggggtattactatacacactacaaaatgtgcagctcctgggagtgtggctat 983
Db 601 TTGGGAAGTGGGATTTACTATACACACTACAAATGTGACAGTCTCTGGGATGTGGCTAT 660
Qy 984 gacttgatgaaacgaacatgctgctggactatacaatggcatacaatgcatactccacacag 1043
Db 661 GACTTGTATGAAACGACAATGCTGCCCTGGGACTATGACAAATGGCATATACTTCCACAG 720
Qy 1044 atgtacactcagagtagtacagcaaatcttagcttcccataacccccacaaagcctatttt 1103
Db 721 ATGTACACTCAGAGAGTACAGCAAAATCTTAGCTTCCCAATAACCCACAAAGCCTATATT 780
Qy 1104 ttatatgtgctatcaagctgttcaatccacactgcaagctcctctggcaggtatttcgaa 1163
Db 781 TTATATATTTGGCTATCAAGCTGTTTCATTTCACCACCTGCAAGCTCTCTGGCAGGTATTCGAA 840
Qy 1164 cactaccgactcattatacaataaacagagagagatgctgcgcactcttctcgtcta 1223
Db 841 CACTACCATTCATTTATCAACATAAACAGGAGGAGATATGTCGCATGCTTTCCTGCTTA 900
Qy 1224 gatgaagcaatcaacaacgtagacattggtcctaaagacttatggtttctataacaacagc 1283
Db 901 GATGAAGCAATCAACAACGTGACATTGGCTCTAAAGACTTATGGTTTCTATAACAACAGC 960
Qy 1284 attatcttactcttcagataaatgtggccagcctacgagaggagagtaactggcct 1343
Db 961 ATTATCATTTACTCTTCAGATAAATGTGGCGAGCCCTACGGCAGGAGGAGTAACCTGGCCT 1020
Qy 1344 ctcaaggtagcaaaagaaacatatattgggaagaggatccgggctgtaggcttttgtcat 1403
Db 1021 CTCAGAGGTAGCAAAAGAAACATATTGGGAAGAGGGATCCGGGCTGTAGGCTTTGTGCTAT 1080
Qy 1404 agccactcttgaaaaacaaggaacagtcgtgtaaggaaacctgtgcacatcaactgag 1463
Db 1081 AGCCCACTTCTGAAAAACAAGGAACAGTGTGTAAGGAACCTTGTGCACATCAGTACTGG 1140
Qy 1464 taccacactctcatttcaactggtcgtgaaagacagatgtatgaggacatttcaactagatggc 1523
Db 1141 TACCCCACTCTCATTTTCATCTGGCTGAAGACAGATTTGATGAGGACATTTCAACTAGATGGC 1200
Qy 1524 tatgatctctggagaccataagttagggtcttgcctcacccccaggttagatattttgcat 1583
Db 1201 TATGATATCTGGGAGACCATTAAGTGGGGTCTTCGCTCACCCCGAGTAGATATTTTGGCAT 1260
Qy 1584 aacattgaccctcatatacaccaagcaaaaaatggctcctggcagcagggctatgggac 1643
Db 1261 AACATTGACCCCATATACACCAAGCAAAAAATGCTCCTCTGGCAGCAGGCTATGGGATC 1320
Qy 1644 tggaaacactgcaatccagtcagccatcagagtcagagcactggaaattgcttacaggaat 1703

```
Db 1321 TGAACACTGCAATCAGTACGCCATCAGAGTGCAGCACTGGAAATTTGCTTACAGAAAT 1380
QY 1704 cctgctcagcagctgggtcccccctcagttctttcagcaaacctggaccgaacggtag 1763
Db 1381 CTTGCTACAGCGACTGGGTCCCCCTCAGTCTTTCAGCAACTGGGACCGAACC GGTTG 1440
QY 1764 cacaatgaacgagatcacctcgtcaactggcaaaagtgtatgcttttcaacatcacagcc 1823
Db 1441 CACAATGAACGGATCACCTTGTCAACTGGCAAAAGTGTATGGCTTTTCAACATCACAGCC 1500
QY 1824 gaccatatgagaggggtggacctatctaacagggtatccaggaaatcgtgaagaagtccta 1883
Db 1501 GACCCATATGAGAGGGTGGACCTATCTAACAGGTATCCAGGAATCGTGAAGAAGTCCTA 1560
QY 1884 cggagctctcacagtccaacaaaactcagtcgcccgcgtcaggtatcccccaagacc 1943
Db 1561 CGAGGCTCTCACAGTTCAACAAAACACTGCAGTCCGGTCCAGGTATCCCCCAAGACCCC 1620
QY 1944 agaagtaacctaggctcaatggaggggtcttagggaccatggtatagagaggaacccaag 2003
Db 1621 ACAAGTAACCTTAGCTCAATGGAGGGGTCTGGGACCATGTATTAAGAGGAAACCAAG 1680
QY 2004 aaaaagaagcgaagcaaaatcagctcagaaaaagcaaaagaaagc-aaaaaaagaag 2062
Db 1681 AAAAAAGAGCCCAAGCAAAAATCAGGCTCAGAAAAAGCAAAAAGCAAAAAGAAAG 1740
QY 2063 aagaacagcagaagcagctcaggttcaacttccattcaggttacttctgtgataa 2122
Db 1741 AAGAAACAGCAGAAAGCAGTCTCAGGTCAACTTCCATTCCAGTTCAGGTGTACTTGTGGATAA 1800

RESULT 3
AC104779/c
LOCUS Homo sapiens chromosome 4 clone RP11-26P13, linear PRI 09-FEB-2002
DEFINITION AC104779
ACCESSION AC104779
VERSION AC104779.4 GI:18642899
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 152861)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 152861)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 152861)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 9, 2002 this sequence version replaced gi:18482303.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: H_NH0026P13
-----
Location/Qualifiers
1. 152861
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
```

```
BASE COUNT 48870 a 27681 c 28505 g 47805 t
ORIGIN
Query Match 66.9%; Score 1515; DB 9; Length 152861;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1529; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 719 aaagtatcagatacacacacggacttcaacattctatcataagacactaccacccaactg 778
Db 10199 AAGGTATCAGATACACACCGGACTTCAACATTTCTATCAAGACCTTACCCACCAACTG 10140
QY 779 ttactctctggacaatgcccacctacccctcagaaactgaagggagtggtgatattcaacgca 838
Db 10139 TTTACTCTGGACAATGCCACCTTACCTCAGAAATGAAGGAGTTGGATATTCAACGCA 10080
QY 839 tatgctcgaaaaatggcacttgggtttttacagaaaaaagaatgcattgccaccacagaagag 898
Db 10079 TATGCTCGAAAAATGGCACTTGGGTTTTTACAGAAAAAAGAAATGCATGCCACCAGAAGAG 10020
QY 899 atttgataacctttttgttcccttttgggaagtggggaattactatacacactacaaatg 958
Db 10019 APTTGATACCTTTTTTGGTTCCCTTTTGGGAAAGTGGGATTACTATACACACTACAAATG 9960
QY 959 tgacagctcgtggatgtgtgctatgacttgtataaaacagacaatgctgctggacta 1018
Db 9959 TGACAGTCTCGGGATGTGGGCTATGACTTGTATGAAAACGACAATGCTGCCTGGACTA 9900
QY 1019 tgacaatggcataatactccacacagatgtacactcagagagtacagcaaatcttagcttc 1078
Db 9899 TGACAAATGGCATATACTCCACACAGATGTACACTCAGAGAGTACAGCAATCTTAGCTTC 9840
QY 1079 ccataacccccacaagcctatatattttatatatttgcctatacgaagtgttcattccaccat 1138
Db 9839 CCATAACCCCCACAAAGCCCTATATTTTATATATTGCTATCAAGCTGTTTCATTTCACCACT 9780
QY 1139 gcaagctcctgcaggtatttcgaacactaccgattccattatcaacataaacacagagag 1198
Db 9779 GCAAGCTCTGCGCAGATATTCGAACACTTACCGATCCATTATCAACATNAACAGGAGAG 9720
QY 1199 atatgctgccatgcttctcgtcttagatgaagcaatcaacaacgtgacattggctctaa 1258
Db 9719 ATATGCTGCCATGCTTCTCTGCTTAGATGAAGCAATCAACAACGTGACATTGGCTCTAAA 9660
QY 1259 gacttatggtttctataacaacacagcattatacttactctcagataatggtggccagcc 1318
Db 9659 GACTTATGCTTCTATAACAACAGCATTTATCATTTTACTCTTTCAGATTAATGGTGGCCAGCC 9600
QY 1319 tacgacagagagagtaaacctgctcctcagaggtagcaaggaaacatatatgggaagagag 1378
Db 9599 TACGGCAGGAGGAGTAACTGGCCCTCTCAGAGGTAGCAAGGAAGAAACATATTGGGAAGGAGG 9540
QY 1379 gatccgggctgtaggcttttgtcctagccccacttctgaaaaaagggaacagtggtgtaa 1438
Db 9539 GATCCGGGCTGTAGGCTTTGTGCATAGCCCACTTCTGAAAAACAAGGGAACAGTGTGTAA 9480
QY 1439 ggaacctgtgaacatcaactcagctggtgtaccccaactctcattcactggtcgaaggacagat 1498
Db 9479 GGAACCTTGTGCACATCACTGACTGGTGTACCCCACTCTCATTTTTCACCTGGCTGAAGAGCAGAT 9420
QY 1499 tgatgagacattcaactagatggctatgatatctgggagaccataaagtggaggtcttcg 1558
Db 9419 TGATGAGGACATTCAACTAGATGGGTATGATATCTGGGAGACCAATAGTGAAGGCTTTCG 9360
QY 1559 ctcaaccccgagtagatattttgtcataacattgacccccatatatacacaagggaacaaatgg 1618
Db 9359 CTCACCCCGAGTAGATATTTTGCATAAATTCACCCCATATACACCAAGGCAAAAAATGG 9300
QY 1619 ctctctgggcagcagcgtatgggatctggaacactcgaactcccaatccagtcagccatcagatgca 1678
Db 9299 CTCCTGGGCAGCAGGCTATGGGATCTGGAACACTGCAATCCAGTTCAGCCATCAGACGTGCA 9240
```

Qy 1679 gcaatggaaattgttacaggaaatcctggctacagagactgggtgtccccctcagctttt 1738
 Db 9239 GCACGTGAAATTTGCTTACAGGAAATCCTGGCTACAGCGACTGGGTCCCCCTCAGCTTT 9180
 Qy 1739 cagcaacctggagccgaacggtggcacaatgaacgataccctcgtaacctgcaaaaag 1798
 Db 9179 CAGCAACCTGGGACCGAACCCTGGCTGCAATGAACGGATCACTTGTCACCTGCAAAAG 9120
 Qy 1799 tgtatggcttttcaacatcacagccgaccccatatgagaggggtgacacctatctaacagta 1858
 Db 9119 TGTATGGCTTTTCAACATCACAGCCGACCCCATATGAGAGGGTGGACCTATCTAACAGTA 9060
 Qy 1859 tccaggaatctgaaagctccctacagaggtctctcacagttcaacaaaactgcagtgcc 1918
 Db 9059 TCCAGGAATCTGTAAGAAGCTTCCTACGGAGGCTCTCACAGTTCACAAAACGTCAGTGCC 9000
 Qy 1919 ggtcaggtatcccccccaagaccacaagtaaccctaggctcgaatgaggggtctagg 1978
 Db 8999 GGTcAGGTATCCCCCAAGACCCCAAGTAACCTTAGGCTCAATGGAGGGGTCTGGGG 8940
 Qy 1979 accatggtatagagagaaacaaagaaagaaagccaaagcaaaaatcaggtctgagaaaaa 2038
 Db 8939 ACCATGTTATAAGAGGAACCAAGAAAGAACGCCAAGCAAAATCAGGCTGAGAAAA 8880
 Qy 2039 gcaaaagaaagc-aaaaaaagaaagaaacacagcagaagcagttcaggttcaacttg 2097
 Db 8879 GCAAAAGAAAAGCAAAAAAGAAAGAAACAGCAGAAAGCAGTCTCAGGTTCAACTG 8820
 Qy 2098 ccattcaggttacttgatgaagcaacaataattcctgttgggttaacttaatca 2157
 Db 8819 CCATTcAGGTGTTACTTGTGGATAGCAACAATATTTCTCTGTGGTAAACTTTAAATCA 8760
 Qy 2158 gtcttattcttctatctgtttcctcaggttaaacaccagcaaatgttgcgcgataatcgcgtg 2217
 Db 8759 GTTCTTATCTTTTCTCTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 8700
 Qy 2218 gcctaagcgtcagcgtgttttctatcgtgtgccac 2252
 Db 8699 GCCTAAGCGTCAGGCTGTTGTTTCTATGCTGTGCCAC 8665

RESULT 4
 AC021342/c
 LOCUS
 DEFINITION Homo sapiens clone RP11-2413, WORKING DRAFT SEQUENCE, 14 unordered
 pieces.
 AC021342
 VERSION AC021342.3 GI:10045513
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 165147)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-2413
 Unpublished
 2 (bases 1 to 165147)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G., Castle,A.,
 Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lander,E., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPeeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Olivat,T.M., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

TITLE JOURNAL COMMENT

Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 9, 2000 this sequence version replaced gi:6899755.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4477
 Center clone name: 24_I_3

----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 156719 bases at least Q40
 Consensus quality: 160377 bases at least Q30
 Consensus quality: 162195 bases at least Q20
 Insert size: 170000; agarose-fp
 Insert size: 163847; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; agarose-fp
 Quality coverage: 4.6 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 867: contig of 867 bp in length
 968 967: gap of 100 bp
 968 2133: contig of 1166 bp in length
 2134 2233: gap of 100 bp
 2234 3987: contig of 1754 bp in length
 3988 4087: gap of 100 bp
 4088 6971: contig of 2884 bp in length
 6972 7071: gap of 100 bp
 7072 10879: contig of 3808 bp in length
 10880 10979: gap of 100 bp
 10980 15646: contig of 4667 bp in length
 15647 15746: gap of 100 bp
 15747 25208: contig of 9462 bp in length
 25209 25308: gap of 100 bp
 25309 33458: contig of 8150 bp in length
 33459 33558: gap of 100 bp
 33559 46617: contig of 13059 bp in length
 46618 46717: gap of 100 bp
 46718 60565: contig of 13848 bp in length
 60566 60665: gap of 100 bp
 60666 92690: contig of 32025 bp in length
 92691 92790: gap of 100 bp
 92791 122716: contig of 29926 bp in length
 122717 122816: gap of 100 bp
 122817 149492: contig of 26676 bp in length
 149493 149592: gap of 100 bp
 149593 165147: contig of 15555 bp in length.

FEATURES
 Location/Qualifiers
 1..165147
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-2413"
 /clone_lib="RP11-11 Human Male BAC"
 1..867
 /note="assembly_fragment
 clone_end:sp6"

```
misc_feature      vector_side:left
968..2133
/note="assembly_fragment"
2234..3987
/note="assembly_fragment"
4088..6971
/note="assembly_fragment"
7072..10879
/note="assembly_fragment"
10980..15646
/note="assembly_fragment"
15747..25208
/note="assembly_fragment"
25309..33458
/note="assembly_fragment"
33559..46617
/note="assembly_fragment"
46718..60565
/note="assembly_fragment"
60666..92690
/note="assembly_fragment"
92791..122716
/note="assembly_fragment"
122817..149492
/note="assembly_fragment"
149593..165147
/note="assembly_fragment
clone_end:17
vector_side:right"
BASE COUNT      52130 a 29805 c 30056 g 51853 t 1303 others
ORIGIN

Query Match      66.9%: Score 1515; DB 2; Length 165147;
Best Local Similarity 99.6%: Pred. No. 0;
Matches 1529; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 719 aaagtacagatcacacacggaacttcaactctatcataagacactcaccacccaactg 778
Db 101883 AGGGTATCAGATACACACCGGACTTCACATCTCATATAGACCTACCCACCCCACTG 101824

Qy 779 ttacacctggacaatgcacacctacacctcagaacactgaagggttgatattcaacgca 838
Db 101823 TTATCCTCTGGCAATGCCACCTACCTACCAAACTGAAGGAGGTGGATATTCAACGCA 101764

Qy 839 tatgtcgaataatgcacttgggtttttacagaaagaatgcattgccaccagaagagg 898
Db 101763 TATGTCGGAATAATGGCACTTGGGTTTTTACAGAAAAGAAATGCATGCCACCAGGAAGAGG 101704

Qy 899 atttgataaccttttgggttcccttttgggaagtgggtgattactatcacactacaaatg 958
Db 101703 ATTTGATACCTTTTGTGCTTCCCTTTTGGGAAGTGGGATTACTATACACTACAAATG 101644

Qy 959 tgacagtctcgggagtgtgtggtctgaacttgatgaacagacaatgtgctcgggacta 1018
Db 101643 TGACAGTCTCGGAGTGTGGCTATGACTTGTATGAAAACGACAATGCTGCTGGGACTA 101584

Qy 1019 tgacaatggcatatactccacagaatatacactcagagatagacagaacttagcttc 1078
Db 101583 TGCAATGGCATATACTCCACACATGTACACTCAGAGGTACAGCAAAATCTTAGCTTC 101524

Qy 1079 ccaatacccccaagaacctatatttttatattgcctatcaagctgttcattccacct 1138
Db 101523 CCATAACCCACAAAGCCTATATTTTATATATATGTCCTATCAAGCTGTTCATTCCACT 101464

Qy 1139 goaagctcctggcaggtatttgcgaacatacogatccattatcaacataaacagaggag 1198
Db 101463 GCAAGCTCTGGCAGGTATTTTCGAACACTACCGATCCATTATCAACATAAAACAGGAGGAG 101404

Qy 1199 atatgctgccatgcttccctcttagatgaagcaatcaacacgtgacattggctctaaa 1258
Db 101403 ATATGCTGCCATGCTTTCTCTTAGATGAAGCAATCAACACGTTGACATGGCTCTAAA 101344
```

```
Qy 1259 gacttatgtttctataacaacagcattatcatttactcttcagataaattggtgcagacc 1318
Db 101343 GACTTATGTTTCTATATAACAACAGCATTTATCATTTACTTTCAGATATATGTTGGCCAGCC 101284

Qy 1319 tacggcaggaggagtaactggcctctcagaggtgagcaaaagaaacatatattgggaaggagg 1378
Db 101283 TAGGCGCAGGAGGAGTAACGTGGCTCTTCAGAGGTAGCAAAAGGAACATATTGGGAAGGAGG 101224

Qy 1379 gatccggcgttaggtctttgtgcataagcccactcttgaataaaggaaggaacagtgtgtaa 1438
Db 101223 GATCCGGCTGTAGGCTTTGTGCATAGCCCACTTCTGAAAAACAAGGGAACAGTGTGTAA 101164

Qy 1439 ggaacctgtcacatcactgactgtaccactctcatttcaactggctggaaggacagat 1498
Db 101163 GGAACCTTGTGCACATCACTGACTGGTACCCCACTCTCATTTTCACCTGGCTGAAGGACAGAT 101104

Qy 1499 tgaatgagacattcaactagatggctatgatatctgtggagacccaataagtgaagggtcttcg 1558
Db 101103 TGATGAGGACATTCAACTAGATGGCTATGATATCTGGGAGACCATAAGTGAAGGCTCTTCG 101044

Qy 1559 ctcaaccggagtagatatatttgataaacaattgaacccaataatacacaagggaacaaatgg 1618
Db 101043 CTCACCCCGAGTAGATATTTTGCATAAACATTGACCCCACTATACACCAAGGCAAAAATGG 100984

Qy 1619 ctctctggcagcaggtctatgggatctggaacactgcaatccagtcagccatcagagtga 1678
Db 100983 CTCCTGGCGCAGCAGGCTATGGGATCTGGAACACTGCATCCACTCAGCCATCAGAGTGA 100924

Qy 1679 gcaatggaaattgcttacagaaaatccctggctacagcagactgggttccccctcagctcttt 1738
Db 100923 GCATCGGAAATTCCTTACAGGAAATCCTTGGCTACAGCGACTGGTCCGCCCTCAGTCTTT 100864

Qy 1739 cagaacctgggacgacccgggtggcacaatgaacggatcaacctcgtcaactggcaaaag 1798
Db 100863 CAGCAACCTGGGACCGAACCGGTGGCAATGAAGCGATCACTTGTCAACTGGCAAAAG 100804

Qy 1799 tgaatggcttttcaacatcacagccgaccccatatgagaggtggagacctatctaaacagga 1858
Db 100803 TGTATGGCTTTTCAACATCACAGCCGACCCCATATGAGAGGTGGGACCTATCTTAACAGGTA 100744

Qy 1859 tccaggaaatcgtgaagaagctcctacggaggtctcagagttcacaacaaactgcagtgcc 1918
Db 100743 TCCAGGAATCGTGAAGAAAGCTCCTACGGAGGCTCTCACAGTTTCAACAAAACTGCAGTGCC 100684

Qy 1919 ggtcaggtatcccccaagaagcccccagagtaaacctcaggtcgaatgaggggtcagagg 1978
Db 100683 GGTCAAGGTATCCCCCAAGAGACCCCAAGTAACCTAGGCTCAATGAGGGGTCTGGGG 100624

Qy 1979 accatggtatagaggaagaaaccaagaaaaaagaagccaaagcaaaatcagggtcgagaaaaa 2038
Db 100623 ACCATGGTATAAAGAGGAACCAACAGAAAAAGAGCCAAAGCAAAANTCAGGCTGAGAAAAA 100564

Qy 2039 gcaaaagaaaagc-aaaaaaagaagaagaacacagcagagaagcagtcctcaggttcaacttg 2097
Db 100563 GCAAAAGAAAAAGCAAAAAAAGAAAGAACACAGCAGAAAGCAGTCTCAGGTTTCAACTTG 100504

Qy 2098 ccattcaggttacttgtgataagcacaataatttctgtgttgttaaacctttaatca 2157
Db 100503 CCATTCAAGGTGTACTTGTGGATAAGCAACAATATTCTGTGTGTAAACTTTAAATCA 100444

Qy 2158 gttctattctttcaactgtttcctcaggttaaacacagcaaatgttggctcgataataatcgctg 2217
Db 100443 GTTCTTATCTTTTCATCTCTTTCTTAGGTAACACAGCAAAATTTGGCTCGATATAATATCGCTG 100384

Qy 2218 gcctaagcgtcaggcttgttttcatgctgtgtgccac 2252
Db 100383 GCCTAAGCGTCAGGCTTGTCTTTCATGCTGTGCCAC 100349

RESULT 5
AC091322/c AC091322 bp DNA linear HTG 11-JUN-2001
LOCUS Mus musculus clone RP23-60D3, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION
```



```

pieces.
AC091322
AC091322.2 GI:14336533
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Graniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 232951)
Birken,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-60D3
Unpublished
2 (bases 1 to 232951)
Birken,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierrel,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., ROY,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-APR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 11, 2001 this sequence version replaced gi:13625486.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L13252
Center clone name: 60_D_3
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 231176 bases at least Q40
Consensus quality: 231965 bases at least Q30
Consensus quality: 232319 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 232551; sum-of-contigs
Quality coverage: 10.7 in Q20 bases; agarose-fp
Quality coverage: 9.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 53: contig of 53 bp in length
*
* 54 153: gap of 100 bp
*
* 154 2590: contig of 2437 bp in length
*
* 2591 2690: gap of 100 bp

```

```

FEATURES
    source
        2691 11774: contig of 9084 bp in length
        * 11775 11874: gap of 100 bp
        * 11875 100524: contig of 88650 bp in length
        * 100525 100624: gap of 100 bp
        * 100625 232951: contig of 132327 bp in length.
    Location/Qualifiers
        1..232951
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone="RP23-60D3"
        /clone_lib="RPC1-23 Female Mouse BAC"
        1..53
        /note="assembly_fragment"
        clone_end:SP6
        vector_side:left"
        misc_feature
            154..2590
            /note="assembly_fragment"
        misc_feature
            2691..11774
            /note="assembly_fragment"
        misc_feature
            11875..100524
            /note="assembly_fragment"
        misc_feature
            100625..232951
            /note="assembly_fragment"
            clone_end:T7
            vector_side:right"
BASE COUNT 70795 a 45474 c 46842 g 69434 t 406 others
ORIGIN

Query Match 50.1%; Score 1134.2; DB 2; Length 232951;
Best Local Similarity 84.7%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 218; Indels 18; Gaps 2;

QY 712 ttactgaaagtatcacatcacacccggacttaacattctatcataagaacctaccacaac 771
Db 147754 TTCTCTAAAGGTATCAGATACACACGGGACTTCAGCATTTCTATCATTAGACCTACCCCAAC 147695

QY 772 ccaactgttacctctctggacaatgccacctaccctcagaaactgaagaggttgatatt 831
Db 147694 CCAACTGCTACTCTTGACAAATGCACACCTTACCTCAGAACCTAAGAGGTCGGCTATT 147635

QY 832 caacgcatatggtcggaaaaatggcacttgggttttttacagaaaaaatgcataccccca 891
Db 147634 CAACCCATATGGTTGGGAAGTGGCACTTGGGATTTTACAGAAAAAGACTGCATGCCACCA 147575

QY 892 gaacgagatttgataccttttttggttcccttttgggaagtggggttactatcacact 951
Db 147574 AGAGAGGATTGTATACCTTTTTTGGCTCTCTTTTGGGAAGTGGTATTTATATACACACT 147515

QY 952 acaaatgtgcagtcctctgggatgtggtctatgacttqtatgaaaacgacaatgctgct 1011
Db 147514 ACAANTGTGACAGTCTCTGGGGTGTGGCTATGACTTGTAGAAAACGACAAATGCAGCTT 147455

QY 1012 gggactatgacaatggcatactactccacacagatgtacactcagagatgacagaaatct 1071
Db 147454 GGGACTATGACAATGGCATCTATTCAACACAGATGTACACGACAGAGTCGACGAAATCT 147395

QY 1072 tagcttccccataacccccacaaagcctatatattttatatattgcctatcaagctgttca 1131
Db 147394 TAGCCACCATGATCCCAACAAACCTCTATTATATATGTGGCTTACCAAGCTGTGCCACT 147335

QY 1132 caccattcaagctcctctggcaggtatttcgaaactaccgatccattatcaacataaaca 1191
Db 147334 CACCACCTTCAGAGCCCTGGCAGGTACTTTGAACACTACAGATCCATTATCAACATAAACA 147275

QY 1192 ggaaggagatagtgcgcaatgctttcctctcttagatgaagcaatcaacaacgtgcattgg 1251
Db 147274 GGAGGAGGTACGAGCCCATGCTCTCTCTGCTGGATGAAGCAATCCACAATGTGACCCCTGG 147215

QY 1252 ctctaaagactattggtttcttataaacaacagcattatcatttactcttcagataatggtg 1311
Db 147214 CCCTAAAGAGGTATGGTTTCTATATAACAATAGCATTTATCTTACTCTCAGATAATGGTG 147155

```


Db	319	CTCAGCTTTTCAGCAACCTGGACCGAACCCTGGCGCAATGAACGGATCACCTTGTCAA	378
Qy	1789	ctggcaaaagtgtatggcttttcaacatcacagccgacccatatgagaggtgacccat	1848
Db	379	CTGGCAAAAGTGTATGGCTTTTCACATCACAGCCGACCCATATAGAGGGTGGACCTAT	438
Qy	1849	ctaacaggtatccagggaatcgtgaagaagctctctacgaggtctcacagttcaacaaaa	1908
Db	439	CTAACAGGTATCCAGGAATCGTGAAGAAGCTCCTACGGAGGCTCTCACAGTTCAACAAAA	498
Qy	1909	ctgcaagtccggtcaaggtatcccccaagaccaccagaaagtaacctaggctcaatgag	1968
Db	499	CTGCAGTCCCGGTCAAGTATCCCCCAAGACCCCAAGTAACCTAGGCTCAATGGAG	558
Qy	1969	gggtctaggaccatggtatagagaggaaaccaagaaaaagcaagcgaagcaaaatcagg	2028
Db	559	GGGTCTGGGACCATGGTATTAAGAGGAAACCAAGAAAAAGCAAGCAACCAAAATCAGG	618
Qy	2029	ctgagaaaaaagcaaaagaaagc-aaaaaagaaagaaagacagcagaaagcagctctcag	2087
Db	619	CTGAGAAAAAGCAAAAGAAAGCAAAAAAGAAAGAAAGCAAGCAAGCAAGCTCTCAG	678
Qy	2088	gttcaacttgccattcaggttacttcttgatgaagcaacaatatctctgttggttaa	2147
Db	679	GTTCAACTTGCCTATTCAGGTGTACTTGTGGATAGCAACAATATTTCTGTTGGTTAA	738
Qy	2148	actttaacagttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct	2207
Db	739	ACTTTAATCAGTCTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	798
Qy	2208	aatatcgctgctgaagcagctgagctgttcttcttcttcttcttcttcttcttcttcttct	2252
Db	799	AATATCGCTGGCTAAGCGTCAGGCTTGTCTTTCATGCTGTGCCAC	843
RESULT	7		
AC013692/c			
LOCUS		157043 bp	DNA linear HTG 07-DEC-2000
DEFINITION		Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***	45
ACCESSION	AC013692		
VERSION	AC013692.5	GI:11597077	
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	1 (bases 1 to 157043)		
REFERENCE	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,		
AUTHORS	Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,		
	Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,		
	Cooke, P., DeArellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,		
	Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,		
	Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,		
	Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,		
	Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,		
	McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,		
	Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,		
	Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,		
	Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,		
	Tesfaye, S., Tirrell, A., Vassiliou, H., Vo, A., Wheeler, J., Wu, X.,		
	Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome		
	Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	On Dec 7, 2000 this sequence version replaced gi:10280848.		
	All repeats were identified using RepeatMasker:		
	Smit, A.F.A. & Green, P. (1996-1997)		

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4053
Center clone name: 21_L_8

* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 924: contig of 924 bp in length
* 925 1024: gap of 100 bp
* 1025 1800: contig of 776 bp in length
* 1801 1900: gap of 100 bp
* 1901 2800: contig of 900 bp in length
* 2801 2900: gap of 100 bp
* 2901 3628: contig of 728 bp in length
* 3629 3728: gap of 100 bp
* 3729 4582: contig of 854 bp in length
* 4583 4682: gap of 100 bp
* 4683 5763: contig of 1081 bp in length
* 5764 5863: gap of 100 bp
* 5864 6992: contig of 1129 bp in length
* 6993 7092: gap of 100 bp
* 7093 8691: contig of 1599 bp in length
* 8692 8791: gap of 100 bp
* 8792 10237: contig of 1446 bp in length
* 10238 10337: gap of 100 bp
* 10338 12115: contig of 1778 bp in length
* 12116 12215: gap of 100 bp
* 12216 14480: contig of 2265 bp in length
* 14481 14580: gap of 100 bp
* 14581 16200: contig of 1620 bp in length
* 16201 16300: gap of 100 bp
* 16301 17801: contig of 1501 bp in length
* 17802 17901: gap of 100 bp
* 17902 19560: contig of 1659 bp in length
* 19561 19660: gap of 100 bp
* 19661 21063: contig of 1401 bp in length
* 21062 21161: gap of 100 bp
* 21162 22798: contig of 1637 bp in length
* 22799 22898: gap of 100 bp
* 22899 25224: contig of 2326 bp in length
* 25225 25324: gap of 100 bp
* 25325 27369: contig of 2045 bp in length
* 27370 27469: gap of 100 bp
* 27470 30262: contig of 2793 bp in length
* 30263 30362: gap of 100 bp
* 30363 32450: contig of 2088 bp in length
* 32451 32550: gap of 100 bp
* 32551 34415: contig of 1865 bp in length
* 34416 34515: gap of 100 bp
* 34516 37321: contig of 2806 bp in length
* 37322 37421: gap of 100 bp
* 37422 40476: contig of 3055 bp in length
* 40477 40576: gap of 100 bp
* 40577 50235: contig of 9659 bp in length
* 50236 50335: gap of 100 bp
* 50336 53457: contig of 3122 bp in length
* 53458 53557: gap of 100 bp
* 53558 56770: contig of 3213 bp in length
* 56771 56870: gap of 100 bp
* 56871 59626: contig of 2756 bp in length
* 59627 59726: gap of 100 bp
* 59727 62247: contig of 2521 bp in length


```
Qy 1597 tataccaagaagcaaaatggctccttgccagcaggctatggatctggaacactgcaa 1656
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1517 TCTACACCATGCCAGCATGGCTCCCTGGAGGGCGGCTTTGGCATCTGGAACACCGCG 1576

Qy 1657 tccagtcagccatcagagtcagcagcactggaattcttacaggaaatcctggctacagcg 1716
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1577 TGCAGGCTGCCATCCGCTGGGTGAGTGGAGCTGCTGACAGGAGACCCCGGCTATGGCG 1636

Qy 1717 actgggtcccccctcagctctttcagcaacctgggaaccagcgggtggcaccaatgaacgga 1776
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1637 ATTGGATCCACCGCAGACATGGCCACCTTCCCGGGTAGTGTGGTGAACCTGGAACGAA 1696

Qy 1777 tcaactcgtcaactggcaaaagtgtatggcttttcaacatcacagccagcccatatgaga 1836
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1697 TGGCC- --AGTGTGCGCCAGGCGGCTGTGGCTCTTCAACATCAGTCTGACCCCTATGAAAC 1753

Qy 1837 ggggtggaacctatcaacaggttatccaggaatcgtgaagaagctcctcaggggctctcac 1896
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1754 GGGAGGACCTGGCTGGCCAGCGGCGCTGATGTGTCGCCACCCCTGCTGGCTCGCCTGGCGC 1813

Qy 1897 agttcaacaacactcagtcgctcaggtcaggtatcccccaagaccccaagaagtaacctta 1956
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1814 AATATAACCGCACACCATCCCGGTACGCTACCCAGCTGAGAACCCCGGGGCTCATCCTG 1873

Qy 1957 ggcctcaatggagggtctctaggaccactgtatagagagaaacccaagaaaaagaccaa 2016
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1874 ACTTTAATGGGGGCTCTTGGGGGCCCTTGGGCCAGTGATGAGGAAGAGGAGGAAG 1933

Qy 2017 gca 2019
|||
Db 1934 GGA 1936

RESULT 11
LOCUS AX206972 1710 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 13 from Patent WO0155411.
ACCESSION AX206972
VERSION AX206972.1 GI:15394728
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1710)
AUTHORS Glucksmann,M.A., Williamson,M., Rudolph-Owen,L.A. and Tsai,F.Y.
TITLE Human sulfatases
JOURNAL Patent: WO 0155411-A 13 02-AUG-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1..1710
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 340 a 554 c 500 g 316 t
ORIGIN

Query Match 26.2%; Score 594.4; DB 6; Length 1710;
Best Local Similarity 62.2%; Pred. No. 1.7e-167;
Matches 1007; Conservative 0; Mismatches 596; Indels 15; Gaps 4;

Qy 402 atgggggcgcgtggcaggattctggatcctctgcctcctcaacttatggttacctgtcctgg 461
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATGCACACCCCTACACGGCTCTCTCTGTGTCAGCCCTGCTCAGCTTCGGCTACCTGTCCCTGG 60

Qy 462 ggcacggccttagaagaggagaaagggccttactagctcaagctggagagaaacta 521
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GACTGGGCC-----AAGCCGAGCTTCTGTGGCCGACGGGCGGGGAGGTGGCGA--G 111

Qy 522 gagccagcaacaactccactccagcccatctcatttcattcctcagcgatgatcag 581
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112 CAGCCCTCGGGCGCTCCCGCCAGGCTCCCCACATCATCTTCTCATCTCTCAGGACGACCAA 171
```

```
Qy 582 ggattagagatgtgggttaccacggatcttgagattaaaacaccttactcttgacaagctc 641
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 172 GGCTACCACGAGCTGGGCTACCATGTTTCAGATATCGAGACCCCTACGCTGACAGGCTG 231

Qy 642 gctgccgaagaggttaaaactggagaactactatgtccagcctatttggcacacatccagg 701
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 232 GCGGCCAAGGGGGTCAAGTTGGAGAAATTATTACATCCAGCCCCATCTGCACGCTTCGCGG 291

Qy 702 agtcaagttattactggaaaagtatcagatacacacgggaacttcaactctatcataaaga 761
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 AGCCAGCTCCTCACTGGCAGGTACCATCCACAGGACTCCAGCATTTCCATCATCCGC 351

Qy 762 cctaccccaacccaactgtttacctctggacaatgcacacctcactcagaacactgaagag 821
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 352 CCACAGCAGCCCACTGCTGCCCTGGACCAGGTGACACTGCCACAGAGAGTGCAGGAG 411

Qy 822 gtggatattcaacacgcatatgctcggaataatggcacttgggtttttcagaaaaaagtc 881
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 412 GCAGGTTATTCCACCCATATGTTGGCAAGTGGCACCTTTCCTGGGCTCGCTCACGGGAAGGTGT 471

Qy 882 atgcccacagaagaagatttgataaccttttttggcttcccttttgggaagtgggattac 941
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 CTGCCCAACCCGCTGGGGGCTTCGACACCTTTCCTGGGCTCGCTCACGGGAAGGTGT 531

Qy 942 tatcacactacaatatgtgacagtcctgggatgtgtggctatgacttgtatgaacacgac 1001
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 TACACCTATGACAACTGTGATGGCCAGCGGTGTGGGCTTCGAGCTGCACGAGGGTGAG 591

Qy 1002 aatgctcgtgggactatgacaatggcctatctcactcacacagatgtacactcagagata 1061
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 AATGTGGGCTGGG---GGCTCAGCGGCCAGTACTCCACTATGCTTTACGCCACGGCGGC 648

Qy 1062 cagcaaatcttagcttcccataaaccacaaagcctatatattttatatattgacctataa 1121
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AGCCATATCTTGGCCAGCCACAGCCCTCAGCGTCCCTCTTCCTATGTGGCCTTCCAG 708

Qy 1122 gtgttctattcacactcctcctgaggtatttctgaacactaccgataccatattc 1181
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 GCAGTACACACACCCCTGCAGTCCCTCTGTGAGTACCTGTACCGCTACCGCACCATGGGC 768

Qy 1182 acataaacaggaggagatgtgctgcatgcttctcgtcttagatgaagcaatcaacaac 1241
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AATGTGGCCCGCGGAAGTACGCGGCCCATGTGTGAGCTGCATGGATGAGGCTGTGGCAAC 828

Qy 1242 gtgacattggctctaaagacttatgtttctataaacaacagcattatcattactctca 1301
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 ATCACTGGGCCCTCAGCGGTACGGTTTCTACAAACACAGTGTCTATCTTCTCCAGT 888

Qy 1302 gataatggtggccagcctacgagcaggaggagtaactggcctctcagaggttagcaagga 1361
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 GACAATGTTGGCCAGACTTTCCTCGGGGGGCGACAACTGGCCGCTCCGAGGACGAAGGCG 948

Qy 1362 acatatgggaaggaggatccggcctgtaggcttctgtcagtagccactcttgaaaaac 1421
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 ACTTATGGGAAGGTGGCGTGGGGGCCCTAGGCTTTGTCCACAGTCCCTGTCTAAGCGA 1008

Qy 1422 aagggaacagtggttaaggaaacctgtgcacatcactgactggtgacccactctatcca 1481
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1009 AAGCAACGACACAGCCGGGCACCTGATGCACATCATCTGATGTGTACCCGACCCCTGGTGGGT 1068

Qy 1482 ctggctgaagacagattgtgagacattcaactagatggctatgatctctgggagac 1541
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1069 CTGGCAGGTGGTACCACCTCAGCAGCCGATGGCTAGATGGCTACGACGTGTGGCGGCGC 1128

Qy 1542 ataatgagggtcttcgctcaccocagatgatatatttgataaacattgaccccatatac 1601
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1129 ATCAGCAGGGCGGGGCGCTCACACACGAGGAGTCTCTGACAAACATTTGACCACTCTTAC 1188

Qy 1602 accaaggcaaaaaatggctcctggcagcagcaggtctggatctggaacactcaatccag 1661
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1189 AACCATGCCACATGGCTCCCTCGAGGGCGGCTTTGGCATCTGGAACACCCCGCTGCAG 1248
```


	* 11613	11712: gap of unknown length	
*	11713	41121: contig of 29409 bp in length	
*	41122	41221: gap of unknown length	
*	41222	75939: contig of 34718 bp in length	
*	75940	76039: gap of unknown length	
*	76040	181923: contig of 105884 bp in length.	
FEATURES			
source	1. .181923		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="5"		
	/clone="CTB-108B20"		
	/clone_lib="Caltech human BAC library B"		
BASE COUNT	44187 a	44776 c	47833 g
ORIGIN			
	Query Match	21.9%; Score 496.8; DB 2;	Length 181923;
	Best Local Similarity	62.5%; Pred. No. 1.7e-137;	
	Matches 812; Conservative	0; Mismatches 482;	Indels 6; Gaps
QY	720	aagtcacgatacacaccggactcaacattctatcataagaacctaccaccccaactgt	779
Db	101678	AGGTACCATCCACACAGACTCCAGCAATTCATCCGCCACGACGACCCAACTGC	101737
QY	780	ttaacctctggacaatgccaccctaccctcagaactgaaggagtgttgatatcaacgcgt	839
Db	101738	CTGCCCTTGACCAGGTGACACTGCTCCACAGAGTCTCAGGAGCGAGTTATTCACCCAT	101797
QY	840	atggtcggaataatggcaacttgggttttttacagaaagaatgatccccaccagaagaga	899
Db	101798	ATGGTGGGGAAGTGGCACCTGGGCTTCTACCGAAGAGTGTGTGCCACCGTCGGGC	101857
QY	900	tttgtaccttttttgttgcctttttgggaagtgggattactatcacactacaaaatgt	959
Db	101858	TTCGACACCTTCTGGGCTGCTTCACGGGCAATGTGGACTATTACACCTATGACAAC	101917
QY	960	gacagtctctgggaatgtgtggtcatgacttgtatgaaacgcacaatctgcctgggaact	1019
Db	101918	GATGGCCAGGCGTGTGGGCTTCGACCTTCGACGAGGTGAGNATGTGCCCTGGG--	101974
QY	1020	gacaatggcatatactctcacacagatgtacactcagagagtagacaaatcttagctccc	1079
Db	101975	CTCAGCGGCCAGTACTCCACTATGCTTTATGCCACGCGCAGCCATATCTGGCCAGC	102034
QY	1080	cataacccccaaaaagcctatatatttatatttgctctatcaagctgttcattcaccactg	1139
Db	102035	CACAGGCCCTCAGGCTGCCCTCTTCCCTCTATGTGGCTTCACGSCAGTACACACCCCTG	102094
QY	1140	caagctctctggcaggtatttcgaacactaccgatccattatcaaatataaacaggaggaga	1199
Db	102095	CAGTCCCCTGTGTAGTACCTGTACCCGCTACCGACCATGGGCANTGTGCCCGCGGGAAG	102154
QY	1200	tatgtgccaatgctttctgcttagatgaagcaatcaaacagctgcacattggctctaaag	1259
Db	102155	TACGCGCCATGTGACCTGCATGGATAGGCTGTGCGCAACATCACCTGGGCCCTCAG	102214
QY	1260	acttatggtttctatacaaacagcatataattactctctcagataaatggtggcagcct	1319
Db	102215	CGCTACGGTTTCTACAAACACAGTGTCACTATCTTCAGTAGTACAAATGTGTGCCAGACT	102274
QY	1320	acggcaggagggaactggcctctcaaggttagcaagaacatatitgggaaggagg	1379
Db	102275	TTCTGGGGGGCAGCAACTGGCCGCTCCGAGGACGCAAGGGCACTTTATGGGAAGGTGGC	102334
QY	1380	atccggctgttagccttgtgcatagcccaactctgaataaacaggaggaacagtggtgaag	1439
Db	102335	GTGGGGGCTTAGGCTTGTCCACAGTCCCTGCTCAGCGAAGCAACGACAGCCGG	102394
QY	1440	gaacctgtgcacatcactgactggttaccocactctcatttcactggctgaagagcagatt	1499
Db	102395	GCACGTATGCACATCACTACTGTTACTGGTACCCGACCCCTGGTGGTCTGGCAGGTGTAC	102454

Db 1074 TTCATGGGTGGGTCTGCTGAGAGCAGCTGAAACGGAGTGGAGCTCTCAGTAAAG 1133
Qy 1441 aacctgtgcacactcaactgactggtaccacactctcattctcactgactgagagacagattg 1500
Db 1134 GATTGATTCAGCTTCTGATGGTCCCTACTTAGTAACACTAGCTGGTGGAAATTGA 1193
Qy 1501 atgaggacattcaactagatggctatgatatctgtggagaccataagtggaggtcttcgct 1560
Db 1194 ATGGAACCAAGCCATTGGATGGTTTAAACCAATGGGATACAACTAGCAGCAACGAGACTCCTT 1253
Qy 1561 caccctgagtagatatttgcataaacattgacccataatacaccagaagcaaaaatggct 1620
Db 1254 CGCCGAGGGAATCCTCTCCACAACTGACATCTCTGTACCCCA--AAAAGGTGTGCC 1311
Qy 1621 cctgggacagcaggctatgggtctggaacactgcaactcagtcagccatcagagtgcaagc 1680
Db 1312 CCGTACAGCA-----ACACTGGGACACGAGGGTCAAGGCAGCTATCAGAGTTGGGG 1364
Qy 1681 actggaaattgcttacaggaatactggtctacagcagctgggtcccccctcagttcttca 1740
Db 1365 ACTACAAGCTGATTACCGGGACCCAGGCAATGGCAGCTGGGTGCTCCACCTGATGGC- 1423
Qy 1741 gcaacctgggaccgaacccggtggcaccaatgaacggatcacctctgtaactggcgaagaagt 1800
Db 1424 -----CACCTGTATTTTGTACCTGAAATCCAAAGATCGGCTCGRAAAACG 1469
Qy 1801 tatggttttcaacatacacagccaccatagagaggtgagacctatctaacagggtatc 1860
Db 1470 TGTGCTGTTTAACTACACCGCGACCCAAAGACATAAATGATGTGTCAGTAGAGAAAC 1529
Qy 1861 caggaatcgtgaagaagctctacggaggtctctcacagttcaacaaaactgcagtgccgg 1920
Db 1530 CACTTGAGGTCTGAGACTGTGCAAACTACTGGTCCAGTTTAAACATACAGCAGTGCCAC 1589
Qy 1921 tcaggatccccccaaagaccccgagaagtaacccttaggtcctaagtggaggtctaggagac 1980
Db 1590 CTAGATACCCCGCACAGACCCAGGTGCGCCAGCTCTGCATGGTATGCTCTGGGGAC 1649
Qy 1981 catggtga 1987
Db 1650 CGTGGGA 1656

RESULT 14
S48472 1939 bp mRNA linear MAM 08-MAY-1993
LOCUS arylsulfatase B [cats, liver, mRNA, 1939 nt].
DEFINITION S48472
ACCESSION S48472
VERSION S48472.1 GI:258855
KEYWORDS
SOURCE
ORGANISM
cat liver.
Felis catus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
1 (bases 1 to 1939)
AUTHORS Jackson,C.E., Yuhki,N., Desnick,R.J., Haskins,M.E., O'Brien,S.J.
and Schuchman,E.H.
TITLE
Feline arylsulfatase B (ARSB): isolation and expression of the
cDNA, comparison with human ARSB, and gene localization to feline
chromosome A1
JOURNAL
MEDLINE Genomics 14 (2), 403-411 (1992)
REMARK 93052342
GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 117975] from the original journal article.
This sequence comes from Fig. 2.
Map location: A1.
Location/Qualifiers
1..1939
/organism="Felis catus"
/db_xref="taxon:9685"
1..1939
/gene="aryl-sulfatase B, ARSB"

CDS
4..1611
/gene="aryl-sulfatase B, ARSB"
/note="This sequence comes from Fig. 2; ARSB"
/codon_start=1
/product="aryl-sulfatase B"
/protein_id="AAB23941.1"
/db_xref="GI:258856"
/translation="MGRRGASLPRGSPRRPLPGLVLLRLLRLLSPRGAGAD
RPHLFVLAADDLGNVDSPHGSNIRTPHLELAAGVLLDNYTQPLCTPSPRSQLLT
GRYQIHGTGLQHOIIPQPCSCVPLDEKLLPQLKEAGYTHMVGKHLGMYRKECLPT
RRGFDYVGLGSEDEYSHERCALIDSLNVTBCALDFRDEQVATCYKNYSPNIET
EAVNTLITSHPPKPLFLYLALQSVHEPQVPEYLPKPYDFIQDKNHHYYAGWVSLMD
EAVNTLITSHPPKPLFLYLALQSVHEPQVPEYLPKPYDFIQDKNHHYYAGWVSLMD
ASPLLKOKGVKNRELIHISDMPLTLVLRGSLTKGTPKLDGFDVWKTSISGSPSPRKE
LLNIDPNFVIDISPCGKSLAPAKDDSSHPAYLAFNTSLHAAIRHGNNKLLTGYPGCG
CWPPPPSPYNDSAIPSSDPPTKTLPWFDIDQDPERHDLSDRYPHIVEQLLRSLRQFVH
KHSVPVHFPAQDPRCDPKGTGANGPWV"
BASE COUNT 450 a 552 c 503 g 434 t
ORIGIN

Query Match 16.7%; Score 379; DB 4; Length 1939;
Best Local Similarity 56.4%; Pred. No. 1.7e-102;
Matches 830; Conservative 0; Mismatches 575; Indels 66; Gaps 4:

Qy 546 cagcccatctcatttctcattcctcagcagatgtagagatttagagatgtgggtaccac 605
Db 139 CGCGCGCATCTCGTCTCGTGTGGCGGACGACCTGGGCTGGAAGCAGCTGAGCTTCCAC 198
Qy 606 ggaatcgagataaaacacactactcttgacaagctcgtgcgaaggaggttaaacctgag 665
Db 199 GGCTCGAATATCCGACCGCGCACCTGGACGAGCTGGCAGCGCGGGGTGCTCTGGAC 258
Qy 666 aactactatgtccagcctatttgcacaccatccaggagtcagttatttaccaggaaagt 725
Db 259 AACTACTACAGCAACCGCTGTGCACGCATCGCGAGCGAGCTGCTCACCAGCGCGCTAC 318
Qy 726 cagatcacacacggagacttcaacattctctatcaagacctaccaccaacactgtttacct 785
Db 319 CAGATCCACACAGGTTTAAACACACCAATAATCTGCGCCCTGTCTCAGCCCGCTGCTCCT 378
Qy 786 ctggacaatgcacccctaccctcagaacctgaaggaggttgatattcaacgcataatgtc 845
Db 379 CTGGATGAAAACCTCTGCCCCAGCTTCTTAAAGAAGCAGGCTACACTACCATATGTC 438
Qy 846 ggaataatgcactggtgtttttacagaaaagatgcatgccaccacagagagattgcat 905
Db 439 GGAATAATGGCACCTGGGAATGTACCGGAAAGAAATGCTTCCAACCCCGGAGGATTGAT 498
Qy 906 acctttttgttcccttttgggaagtgggtattactatacacactacacaaatgtgacagt 965
Db 499 ACTTACTTTGGATATCTCTTAGGTAGTCAAGATTACTATCCCATGAGCGCTGTGCATTA 558
Qy 966 cctggga-----tggtggtctatgactgtatgataaaacagacactgctgcc 1010
Db 559 ATTGACAGCTCGAATGTGCACACGATGTGCTCTGTGATTTTCGAGATGGGAAACAGGTGCA 618
Qy 1011 tgggactatgacaatggcctatactccacacagatgtacactcagagagatcacagacaaat 1070
Db 619 ACAGATATATAAAAT---ATGTATTTCGACAACATATTTTACTGGAAGAGCTACAGCCCTC 675
Qy 1071 ttgcttcccatcaacccacaaagcctatatatttttataattgtcctatacaagctgttcatt 1130
Db 676 ATAACCTAGCCATCCACCCGAGAGAGCTCTGTTTCTTCTACTTGTCTTTCAGTCTGTCCAC 735
Qy 1131 tcaccactgcaagctcctggaagtgatttcgaacactaccatccatccatatacaacaataac 1190
Db 736 GAACCCCTTCAGGTCCCTCTGAGGAGTACCTGAAACCCCTACGACTTTTATCCAGATAGAAT 795
Qy 1191 aggaggagatagtcgccatgctcttccctgcttagtagaagaacatcaacacagtgacattg 1250
Db 796 AGGCATTACTATGCAGGAATGGTGTCTTATGTGGATGAAGCAGCTGGGAATGTCACAGCA 855

Search completed: July 23, 2002, 17:14:59
Job time: 7805 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	703.8	31.1	817	10	BG680752	602628445	
2	646.2	28.5	683	10	BF307503	601893984	
C 3	603.8	26.6	615	9	AW474222	xs20H05.x	
4	512.4	22.6	625	12	AZ815176	2M0083G13	
5	511.2	22.6	648	9	BB520951	BB520951	
6	506.6	22.4	700	9	BB319509	BB319509	
C 7	393.8	17.4	463	9	AA159953	Z075309.s	
C 8	368.8	16.3	414	10	W81455	z87a04.sl	
9	361.4	15.9	459	10	BF635682	277595.MA	
10	341.8	15.1	547	10	BG661302	1a59e07.Y	
11	333.2	14.7	344	10	W81485	z87a04.r1	
C 12	331.2	14.6	458	10	R43179	vg18d03.sl	
13	307.4	13.6	325	10	U46280	HSU46280.Hu	
14	284.6	12.6	1157	10	BM461300	AGENC007	
15	262.1	11.6	628	10	BE535365	ur47a08.Y	
16	261.4	11.5	625	10	BI104707	602891618	
C 17	245.8	10.8	1032	10	BF979627	602287919	

ORIGIN	AUTHORS	TITLE	JOURNAL	COMMENT
Query Match	31.1%	Score 703.8	DB 10	Length 817


```

Qy 1071 ttatgtcccaataccccacaaagcctatatatttttatattattgctatcatcaagctgttcat 1130
Db 540 TTAGCTTCCCATAAACCCACAAAGCCTATATTTTATATATATGCTATCAAGCTGTTTCAT 599
Qy 1131 tcaccactgaagctcctctgaggtatttccaaacactaccatccattatcaacataaac 1190
Db 600 TCACCACATGCAAGCTCCTGGCAGGTATTTCCGAACACTACCCGATCCATTATCAACATAAC 659
Qy 1191 agggaggagatgctgc 1207
Db 660 AGGAGGAGATGCTGC 676

RESULT 3
AW474222/c
LOCUS
DEFINITION xs20r05.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2770233 3'
similar to SW:ARSB_HUMAN P15848 ARYL SULFATASE B PRECURSOR ;, mRNA
sequence.
ACCESSION AW474222
VERSION AW474222.1 GI:7044328
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 615)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gluco
High quality sequence stop: 435.
FEATURES
source
1..615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2770233"
/clone_lib="NCI-CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"
BASE COUNT 116 a 151 c 153 g 195 t
ORIGIN

Query Match 26.6%; Score 603.8; DB 9; Length 615;
Best Local Similarity 98.9%; Pred. No. 5.3e-106;
Matches 608; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1437 aaggaacctgtgcacatacactgaactggttacccccactctcttcaactggtgaaggacag 1496
Db 615 AAGGAACGTGGTCACATCACCACTAGTGGTACCCCACTCTCATTTCACTGGCTGAAGGACAG 556
Qy 1497 attgatgaggacattcaactgaatgctgatctatctgaggagaccataagtgaaggctctt 1556
Db 555 ATTGATGAGGACATTCACACTAGATGGCTATGATATCTGGGAGACCATTAAGTGAGGGTCTT 496
Qy 1557 cgctacccccgagtagatattttgtcataaacattgacccccatatacccaaggcaaaaaat 1616

```

```

Db 495 CGCTCACCCCGAGTAGATATTTTGCATAACTATGACCCCATATACCAAGCAAAAAAT 436
Qy 1617 ggctctctgggagagagctatgggatactggaaactgaatcccaagcagcatcaagatg 1676
Db 435 GGCTCTCTGGGAGCAGAGGCTATGGGATCTGGAACACTGCAATCCAGTCAGCCATCAGAGTG 376
Qy 1677 cagcactggaatctgttacaggaatcctggctacagcagctgggtccccccctcagct 1736
Db 375 CAGCACTGGAATTCCTTACAGGAAATCTGGCTACAGCGACTGGGTCCCCCTCAGTCT 316
Qy 1737 ttcagcaacctgggagaccgacccggtggcacaatgaacggatcacctcgtcaactatcgcaaa 1796
Db 315 TTCAGCAACCTGGGACCGCAACCGGTGGCACAATGAACGATCACCTTGTCAACTGGCAAA 256
Qy 1797 agtctatgcttttcaaacatcacagccgaccccatatgagaggggtgagacctatctaaacagg 1856
Db 255 AGTGTATGGCTTTTCAACATCACAGCCGACCCATATGAGAGGTGGACCTATCTAACAGG 196
Qy 1857 tatccaggaatcgtgaagaagctcctacgaggtcctcacagttccaacaaactcagtg 1916
Db 195 TATCCAGGAATCGTGAAGAAGCTCCTACGGAGGCTCTCACAGTTCAACAAAACCTCAGTG 136
Qy 1917 ccggtcaggtatcccccaagaccccaagagaagtaaaccttagtctcaatggagggtctag 1976
Db 135 CCGGTCAAGTATCCCCCAAGAGCCCAAGAGTAACCTAGGCTCAATGGAGGGTCTGG 76
Qy 1977 ggaccatggtatagagagaaacccaagaaaaagaaagcaagcaaaaaatcaggctgagaaa 2036
Db 75 GGACCATGTTATAAGAGAGAAACCAAGAAAAGAACCAAGCAAAAATCAGGCTGAGAAA 16
Qy 2037 aagcaaaaagaaaagc 2051
Db 15 AAGCAAAAAGAAAAGC 1

RESULT 4
AZ815176
LOCUS
DEFINITION 2M0083G13f Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUCG2M0083G13 F, DNA sequence.
ACCESSION AZ815176
VERSION AZ815176.1 GI:12985084
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 625)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0083 row: G column: 13
Seq primer: CGTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 625.
FEATURES
source
1..625
/organism="Mus musculus"
/strain="C57Bl/6J"

```

```
/db_xref="taxon:10090"
/clone="UUC2M008G13"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gii4732114[gb]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      179 a  150 c  159 g  137 t
ORIGIN

Query Match      22.6%; Score 512.4; DB 12; Length 625;
Best Local Similarity 90.0%; Pred. NO. 1.8e-88;
Matches 549; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1237 acaacgtgacattgctctaaagacttaattgtttctataaacacagcattatcttact 1296
Db      |||||
QY 16  ACATGTTGACCCCTGAGAGGTATGGTTCTTAACAAATAGCATTAATCAATTACT 75
Db      |||||
QY 1297 ctccagataaattgtgcccagcctacgagcagaggaagtaactggcctctcagaggtagca 1356
Db      |||||
QY 76  COTCAGATAATGTTGGCAGCCACACAGCAGGAGGAAGTAACCTGGCCACTCAGAGGCAGCA 135
QY 1357 aaggaaacattattggaagaggatcccggtctgtaggctttgtgcatagcccaactctga 1416
Db      |||||
QY 136  AAGCAACATATTGGGAAGAGGAGCATCCGGGCACTGGTGGCTTTGTGATAGCCCACTCTAA 195
QY 1417 aaacaagggaacagtgtgtaagggaacctgtgcacactgactgactgtaacccactctca 1476
Db      |||||
QY 196  AAACAAGGGAACGGTATGTAGGAACCTGTGCACATCACCATTGGTACCCCAACTTGA 255
QY 1477 ttctactggctgaaggacagattgtagaggacattcaactagatggctatgatctggtg 1536
Db      |||||
QY 256  TTTCACTGGCTGAAGCAGACAGATTGATGAAGACATTCAGCTAGATGATACGATATCTGGG 315
QY 1537 agacataaagtgaaggttctgctcaaccccgagtagatatttgcataaacattgacccca 1596
Db      |||||
QY 316  AGACCATTAAGCGAAGGCTTCGTTCCACCCGAGTGGATATTTCACCAACATTTGACCCCA 375
QY 1597 tataccaaggaacaaatgctctctggcgagcaggtatggtatggtatggaacactgcaa 1656
Db      |||||
QY 376  TTTACACCAAGCGGAATAATGCTCTCTGGCGACGAGGCTATGGGATCTGGAACACATGCA 435
QY 1657 tccagtcagcattcagtcagcactggaattgtcttacaggaatactctggctacagcg 1716
Db      |||||
QY 436  TCCAGTCGGCCATCCGGGTACAAACACTGGAACACTGGAACACTGCTCAGGCAATCTCTGCTACAGTG 495
QY 1717 actgggtcccccctcagctctttcagcaacctgggacccaacgggtggcacaatgaacgga 1776
Db      |||||
QY 496  ACTGGGTCCCCCTCAGGCTTTTCAGCATCTCTGGGCCCAACCGGTGGCACAATGAAGGA 555
QY 1777 tcaactcgtcaactggcgaagagtgtatggcttttcaacatacacagccgacccatagaga 1836
Db      |||||
QY 556  TTACCTTGTCAACTGCGCAAGAGTATCTGGCTTTTCAACATCAGCGCTGATCCGTATGAGA 615
QY 1837 gngtggacct 1846
```

```
|||||
Db      616 GGGTGGACCT 625

RESULT      5
BB520951
LOCUS      BB520951
DEFINITION  BB520951 RIKEN full-length enriched, 16 days neonate heart Mus
            musculus CDNA clone D830047F08 3', mRNA sequence.
ACCESSION   BB520951
VERSION     BB520951.2 GI:16443460
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 648)
AUTHORS    Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
            Hiramoto,K., Hori,F., Ishi,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
            RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
            Unpublished (2001)
TITLE       On Jul 28, 2000 this sequence version replaced gi:9572409.
JOURNAL     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwaki,S., Inoue,K., Todawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES    cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Location/Qualifiers
            1..648
               /organism="Mus musculus"
               /db_xref="taxon:10090"
               /clone="D830047F08"
               /clone_lib="RIKEN full-length enriched, 16 days neonate
               heart"
               /tissue_type="heart"
```

```
/dev_stage="16 days neonate"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GACAGAGAGCGCGCACTCGAGTGTGTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GACAGAGAGATGTCGAGTTAAATTAATCCCGCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
```

BASE COUNT	179 a	163 c	171 g	135 t
ORIGIN				
Query Match	22.6%	Score 511.2;	DB 9;	Length 648;
Best Local Similarity	87.8%	Preq. No. 3.1e-88;		
Matches 569;	Conservative 0;	Mismatches 78;	Indels 1;	Gaps 1;

```
QY 1346 cagaggtagcaaaagaacatatattgggaaggaggtatccggctgttagcttggcatag 1405
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 1 CAGAGGCAGCTAAGGAACAATATTGGAAGAAGCATCCGGGCAGTGTGTTGTGCATAG 60

QY 1406 ccaactctgaaaaaaa-gggaacagtgtgtaagggaacctgtgcacatcaactactggt 1464
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 61 CCCACTTCTAAAGAACCAAGGGGACGGTATGTGAAGGAACTTGTGCATCATCCGATTGGT 120

QY 1465 acccaactctcatctactggtgaagacagattgatgagacatcaactagatggct 1524
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 121 ACCCAACTTGTATTCTACTGCTGAAGCAGCAGATTGATGAACATTCAGCTAATGGAT 180

QY 1525 atgatattgggagaccataagtgaagggtcttcgctcacccccagtagatatatttgcat 1584
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 181 ACGATATCTGGAGACCATTAAGCAAGGTCTTCGTTACCCCGAGTGATATTTGCACA 240

QY 1585 acattgaccccatatacacaagcaaaaaatgctctctggtggcagcgatctatggatct 1644
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 241 ACATTGACCCCATTTACACCNAAGCGGAAAAATGGCTCTCTGGCGACGAGGCTATGGGATCT 300

QY 1645 ggaacactgcaatccagtcaagccatcagatgcagcactggaatgtgcttacaggaatac 1704
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 301 GGAACACTGCAATCCAGTCGGCCATCCGGGTACAACTGGAACACTGCAACGCAATC 360

QY 1705 ctggctacagcagctgggttccccctcagttcttcagcaacctgggaccgacccggtggc 1764
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 361 CTGGCTACAGTGAAGTGGTCCGCCCTCAGGCTTTTCAGCAATCTGGGCCCAACCCGGTGGC 420

QY 1765 acaatgaacgatcacctctgcaactggcaaaagtgtatgagctttcaacatcacacgcg 1824
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 421 ACAATGAAGGATTACCTTGTCAACTGGCAAGATATCTGGCTTTTCAACATCACGCGTG 480

QY 1825 acccatatgagaggttgagacctatctaacaggtatccaggaatcgtgaagagctctac 1884
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 481 ATCCGTATGAGAGGTTGAGACCTATCCAGCAGGTATCCAGGATCTGTGAAGAGCTGTGTC 540

QY 1885 ggaggtctctcagttcaacaaaaactcagtgccggttcaggtatcccccccaaaacccca 1944
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 541 GGAGACTGTCAATTAACAAGAGCTCCGCTGCTGTCCAGGTACCCCGCAAGGATCCCA 600

QY 1945 gaagtaaccttagctcaatggaaggttcttaggaccatggtatag 1992
      ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 601 GAAGCAACCCCTAGCTCAACGGAGAGTCTGGGGTCCGTGCTGTATAAAG 648
```

RESULT 6

BB319509 700 bp mRNA linear EST 24-OCT-2001

DEFINITION
BB319509 RIKEN full-length enriched, adult male corpora
quadrigenina Mus musculus cDNA clone B230380L03 3', mRNA sequence.
BB319509
VERSION
BB319509.2 GI:16403267
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 700)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanaagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)

TITLE
JOURNAL
COMMENT
On Jul 11, 2000 this sequence version replaced gi:9026544.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

FEATURES
source
1. .700
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="B230380L03"
/clone_lib="RIKEN full-length enriched, adult male corpora
quadrigenina"
/sex="male"
/tissue_type="corpora quadrigenina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in

RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCAGTAATTAATATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda PIC 1."

BASE COUNT	204 a	179 c	182 g	135 t	ORIGIN
Query Match	22.4%	Score 506.6;	DB 9;	Length 700;	
Best Local Similarity	85.4%;	Pred. No. 2.3e-87;			
Matches 599;	Conservative 0;	Mismatches 99;	Indels 3;	Gaps 3;	
QY 1335	aactggcctctcagagtagcaagaggacaacatttgggaaggagggatcctggcctgtaggc	1394			
Db 3	ATCTGCCCCCTCAAGGGCAGCAAGAGGAACATATT-GGAAGGGGGCACCCGGCAGTTGGC	61			
QY 1395	tttgtcatagccaccitctgaaaaaacaaggaagacagtggttaaggaacactgtgcacatc	1454			
Db 62	TTTGGGATAGCCCACTTTTAAACCAAGGACCGGGTATGTAAAGAACTTGT-CACATC	120			
QY 1455	actgactgggtaccaccactctctatttcactggtgaaggacagatgtatgaggacattcaa	1514			
Db 121	CCCGATTGGGACCCCACTTGATTTCCTGGCTGAGGACAGATTGAT-AAACCACTCAG	179			
QY 1515	ctagatggctatgatattctgggagaccataaagtggggtcttcctcaccctcaggtagat	1574			
Db 180	GTAGATGGATACGATATCTGGGAGACCAATAAGCGAAGGTCTTCGTTACCCCGAGTGGAT	239			
QY 1575	attttgcataaattgaccccatatatacacaaggaacaaaatggctcctgggcagcaggc	1634			
Db 240	ATTTTGGCAACAATTGACCCCATTTTACAAAGGCGAAAAATATGGCTCCTGGGCAGCAGGC	299			
QY 1635	tatggatctctggaacactgcattccagtcagccatcagagtgcagcactggaaattgctt	1694			
Db 300	TATGGGATCTGGAACACTGCAATCCAGTCGGCCATCCGGGTACAACACTGGAACTGCTC	359			
QY 1695	acaggaatcctggctacagcagcaatgggtcccccctcagttcttcagcaacctgggaccg	1754			
Db 360	ACAGGGCAATCTGTGCTACAGTACGTGGGTGGTCCCTCCTCAGGGTTTCAGCAATCTGGGCCA	419			
QY 1755	aaccggtggacaaatgaacgcatcactcgttcaactggcaaaagtgtatggcttttcaac	1814			
Db 420	AACCGGTGGCAACATGAAGATTACCTTGTCACTGGCAAGATATCTGGCTTTTCAAC	479			
QY 1815	atcacagccgaccatatgatgagagggtggacactatctaacaggtatccagggaactcgtgaag	1874			
Db 480	ATCAGCGCTGATCCGTATGAGAGGGTGGACCTATCCAGCAGGTATCCAGGCACTGTTGAAG	539			
QY 1875	aagtcctcagagggtctctcaagttcaacaaaaactgcagtcgcggtcaggtatcccccc	1934			
Db 540	AAGTGTGTGGAGAGCTGTCAAAATTCAACAAGACTGCGCTGTCTGTCAGTATCCCCCGC	599			
QY 1935	aaagacccccagaagtacccttagctcaatggagggtcttagggaccattgatatagagag	1994			
Db 600	AAGATCCCAGAACCAACCTTAGCTCAACGGAGAGTCTGGGTCCGTGTGTTAAAGAG	659			
QY 1995	gaaaccaaagaaaaagagccaagcaaaaatcaggctgagaa	2035			
Db 660	GAAACAAGAAAAAGAGTCAAAACAAAACCAAGGCTAAGAA	700			
RESULT 7					
AA159953/c					
LOCUS	AA159953	463 bp	mRNA	linear	EST 16-DEC-1996

DEFINITION					
z075a09.s1 Striatogene pancreas (#937208) Homo sapiens cDNA clone IMAGE:592696 3' similar to SW:ARSB_HUMAN P15848 ARYLSULFATASE B precursor ; , mRNA sequence.					
ACCESSION					
AA159953 GI:173444					
VERSION					
EST.					
KEYWORDS					
human.					
SOURCE					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE					
AUTHORS					
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissov,S., Dietrich,N., DuBueque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.M.					
TITLE					
Generation and analysis of 280,000 human expressed sequence tags					
JOURNAL					
Genome Res. 6 (9), 807-828 (1996)					
MEDLINE					
97044478					
COMMENT					
Contact: Wilton RK					
Washington University School of Medicine					
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108					
Tel: 314 286 1800					
Fax: 314 286 1810					
Email: est@watson.wustl.edu					
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.					
Seq primer: -40W13 fwd. from Amersham					
High quality sequence stop: 369.					
FEATURES					
source					
1..463					
<code>/organism="Homo sapiens"</code>					
<code>/db_xref="GDB:4624082"</code>					
<code>/db_xref="taxon:9606"</code>					
<code>/clone="IMAGE:592696"</code>					
<code>/clone_lib="Stratagene pancreas (#937208)"</code>					
<code>/lab_host="SOIR cells (kanamycin resistant)"</code>					
<code>/note="Organ: pancreas; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pancratic adenocarcinoma cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTGGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTT 3'"</code>					
BASE COUNT 82 a 115 c 115 g 151 t					
ORIGIN					
Query Match 17.4% Score 393.8; DB 9; Length 463;					
Best Local Similarity 97.6% Pred. No. 1.2e-65;					
Matches 442; Conservative 0; Mismatches 7; Indels 4; Gaps 3					
Qy	1603	ccagggaacaaattggtccttgggcagcaggc-tatgggatcttggacactgcataaccag	1661		
Db	452	CCAAGGCAAAAATGGCTCTCTGGCAGCAGGCTTATGGGATCTGGAACACTGCAATCCAG	393		
Qy	1662	tcagccatcacagtgcagcactggaattcctacagaaaacctcgctcacgc-gactg	1720		
Db	392	TCAGCCATFCAGATGCAGACTGGAAATGTGTTACAGAAATCTCGTTACAGCGGACTT	333		
Qy	1721	ggccccctcagttcttcagcaacctgggacc-gaacccggtggcacaatgaacgcatca	1779		
Db	332	GGTCCCCCTCAGTCTTTCACCACTGGGACCGAACC GGTCGCACAATGAACCGATCA	273		
Qy	1780	cctcgtaacctggcaaatgttatggcttttccaacataccagccgccatatgagaggg	1839		
Db	272	CCTTGTCAACTGGCAAAAGTGATGGCTTTTTCACATCATCAGCGCGACCATATGAGAGG	213		
Qy	1840	tggacctatcaacagtattcagaagaatcgtagaagaagctcctcagcaggtctccacagt	1899		
Db	212	TGGACCTATCTAACAGGTATCCAGAATCGTAGAAGAAGT-CTACGAGGGCTCTCACAGT	154		
Qy	1900	tcaacaaactgcagtgccggtcaggtatcccccaagaaccccacagaagttaacccttaqgc	1959		

```

Db 153 TCAACAAATGCGAGTCCGGTCAGGTATCCCCCAAGACCCAGAGTAACCTTAGGC 94
|||||
Qy 1960 tcaatgagggtctaggaccatggtatagaggaggaacacaaagaaagcacaagca 2019
|||||
Db 93 TCAATGAGGGGCTGGGGACCATGTTATTAAGAGAGAAACCAAGAAAAGAACCAAGCA 34
|||||
Qy 2020 aaatcaggctgagaaaaagcaaaagaaaagca 2052
|||||
Db 33 AAAATCAGGCTGAGAAAAGCAAAAAGAAAAGCA 1
|||||

RESULT 8
WB1455/c 414 bp mRNA linear EST 17-OCT-1996
LOCUS z887a04.s1 Soares fetal heart_NbH19W Homo sapiens cDNA clone
DEFINITION IMAGE:347598 3' similar to SW:ARSB_HUMAN P15848 ARYL SULFATASE B
PRECUSOR ;, mRNA sequence.
ACCESSION WB1455
VERSION WB1455.1 GI:1392485
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 414)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 546 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 383.
Location/Qualifiers
1. .414
/organism="Homo sapiens"
/db_xref="GDB:1272973"
/db_xref="taxon:9606"
/clone="IMAGE:347598"
/clone_lib="Soares_fetal_heart_NbH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pR73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTCAAGTGGCGGCGCATCTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pR73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Patima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbH19W."
BASE COUNT 74 a 105 c 105 g 130 t
ORIGIN

Query Match 16.3% Score 368.8; DB 10; Length 414;
Best Local Similarity 97.6%; Pred. No. 7.6e-61;
Matches 406; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

```

```

Qy 1625 ggcagcaggctatgggatctggaacactgcaatccagtcagccatcagatgcagcactg 1684
|||||
Db 413 GGCAGCAGGCTATGGGATCTGGAACACTGCAATCCAGTCCAGTCAGCATCAGAGTCAGCAGCTG 354
|||||
Qy 1685 gaaattccttacaggaatacctggctacagcagactgggtcccccctcaagtctttcagcaa 1744
|||||
Db 353 GAAATTGCTTACAGGAATACTGGCTACAGGACT-GGTCCCCCTCAGTCTTTTCAGCAA 295
|||||
Qy 1745 cctgggaccgaacccgtggtgacaaatgaacgatcacctcgtcaactggcgaagtgatg 1804
|||||
Db 294 CCTGGGACCGAACCGGTGGCACAAATGAACGATCACCTTGTCAACTGGCAAAAGTGATG 235
|||||
Qy 1805 gctttcaacatcacagccgaccccatatgagaggtggacatctctaaacaggtatccagg 1864
|||||
Db 234 GCTTTTCAACATCACAGCCGACCCATATGAGAGGTGGACCTATCTAATCAACAGGTATCCATG 175
|||||
Qy 1865 aatcgtgaagaagctctacgaggtctcacagttctcaacaaactcagtcgcggtcag 1924
|||||
Db 174 AATCGTGAAG-AGAGTCTACGGAGGCTCTCACAGTTCACAAAACTGCAGTGCCGGTCAG 116
|||||
Qy 1925 gtatcccccaagaccccaagaagttaaccctagctcaatgaggggtcttaggaccatg 1984
|||||
Db 115 GTATCCCCCAAAACCCCAAGAGTAACCTAGGCTCAATGGA-GGGTCTGGGACCATG 57
|||||
Qy 1985 gtatagagagaaaccccaagaaagaaagcccaagcaaaatcaggctgagaaaaagc 2040
|||||
Db 56 GTATAAGAGGAAACCAAGAAAGAAAGCAAGCAAGCAAAATCAGGCTGAGAAAAAGC 1
|||||

RESULT 9
BF653682 459 bp mRNA linear EST 25-APR-2001
LOCUS BF653682
DEFINITION 277595 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF653682
VERSION BF653682.1 GI:11918814
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 459)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 66 row: F column: 20
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1. .459
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

```


JOURNAL
COMMENT

Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 546 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 303.

FEATURES
source

Location/Qualifiers
1. .344
/organism="Homo sapiens"
/db_xref="GDB:1272973"
/db_xref="taxon:9606"
/clone="IMAGE:347598"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7m3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGCGCGCATCTTTTTTTTTTTTTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetas as the fetal lung library, Soares fetal lung
NbHL19W."

BASE COUNT 92 a 93 c 82 g 74 t 3 others
ORIGIN

Query Match 14.7%; Score 333.2; DB 10; Length 344;
Best Local Similarity 98.2%; Pred. No. 5.5e-54;
Matches 335; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1551 ggtcttcgctcaaccagtagatatttgcataaacattgaccccatatacacaaggca 1610
|||||
Db 1 GGTCTTCGCTCACCGCAGTAGATATTTCATACATTGACCCCATATACCAAGSCA 60
|||||
QY 1611 aaaaaggctctggcagcagcgtatggatctggaacactgcaatccactcagccatc 1670
|||||
Db 61 AAAAATGGCTCTGGGCAGCAGGCTATGGGATCTGGAACTGCAATCCATCAGCCATC 120
|||||
QY 1671 agagtgcagcactggaattgcttacaggaaatcctggctacagcagctgggtcccccct 1730
|||||
Db 121 AGAGTGCAGCACTGGAAATTCCTTACAGGAATCCTGGCTACAGCGACTGGTCCCCCT 180
|||||
QY 1731 cagcttttcagcaacctgggaccgaacgggtggcacaaatgaacggatcacctcgtcaact 1790
|||||
Db 181 CAGTCTTTTCAGCACTGGGACCGCAACNGTGGCAATGAACGGATCACCTTGTCAACT 240
|||||
QY 1791 ggcaaaagtgtatgcttttcaacatcacagccagccaccatgatgaggggtgagcctatct 1850
|||||
Db 241 GGCAAAAGTGTATGGCTTTTCAACATCACACCCGACCCCATATGAGAGGTGGACCTATCT 300
|||||
QY 1851 aacagggtatccaggaaatcgtaagaagctctctacggaggct 1891
|||||
Db 301 AACAGGTATCCATGANTCGTGAAGANGCTCCTACGGGGCT 341
|||||

RESULT 12
R43179/c
LOCUS
DEFINITION
Yv18403.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:32854 3' similar to SP:ARSB_HUMAN P15848 ARYL SULFATASE B
PRECURSOR ; , mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

R43179
R43179.1 GI:825405
EST.
human.
Homo sapiens

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 458)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT

On May 8, 1995 this sequence version replaced gi:801403.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1275
High quality sequence stops: 384 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1275 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 384.

FEATURES
source

Location/Qualifiers
1. .458
/organism="Homo sapiens"
/db_xref="GDB:405201"
/db_xref="taxon:9606"
/clone="IMAGE:32854"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lfamid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AAGTGAAGAAATTCGCCGCCAGGAATTTTTTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN

74 a 110 c 119 g 152 t 3 others
Query Match 14.6%; Score 331.2; DB 10; Length 458;
Best Local Similarity 92.6%; Pred. No. 1.2e-53;
Matches 401; Conservative 0; Mismatches 26; Indels 6; Gaps 5;

QY 1632 ggctatggatctggaacactgcaatcc--agtacgcatcagagtg-cagcactggaaa 1688
|||||
Db 434 GCCTATGGGATCTGGAACTGCAATCCCATCCAGTCAGCCCATCAGATGCCAGCAGTGAAA 375
|||||
QY 1689 ttgttacagaaat-cctggctacagcg-actgggtcccccctcagctctttcagcaa-c 1745
|||||
Db 374 TTNCTTACAGAAATCCCTGGCTACAGCGAACTGGGTCCCCCTCAGTCTTTTCAGCAACC 315
|||||
QY 1746 ctgggaccgaaccggtggcacaatgaacgcatcactcgtcaactggcacaagtatagg 1805
|||||
Db 314 CTGGACCGAACCGGTGGCACAATGAACGGATCACCTTGTCAACTGGCAAAAGTGTATGG 255
|||||
QY 1806 cttttcaacatcacagccgacccatgatgagagggtggacctatctaaacaggtatccagga 1865
|||||
Db 254 CTTTTCACATCACACCGGACCCCATATGAGAGGTGGACCTATCTTAACAGGTATCCCAGG 195
|||||
QY 1866 atcgtgaagaagctcctcagcaggcctcagagttcaacaaactgcagtcgcggtcagg 1925
|||||

```
Db 194 AATCGTAGANAGNCTACGGAGGCTCTCACAGTTCAACAAAACATGCAGTCCGGTCAGG 135
Qy 1936 tatcccccaagacccagaaagtaacctaggtcaatgaggggtctagggaccatgg 1985
Db 134 TATCCCCCAAGACCCCAAGAAGTAACTAGGCTCAATGGAGGGGTCTGGGACCATGG 75
Qy 1986 tatagagggaaaccaaagaaaaagccaagcaaaatcagggtcgagaaaaagcaaaa 2045
Db 74 TATAAAGAGGAACCAACGAAAAAGAGCAAGCAAAATCAGGCTGACAGAAAAGCAAAAG 15
Qy 2046 aaaaagcaaaaaa 2058
Db 14 AAAAGCAAAAAA 2

RESULT 13
LOCUS U46280 U46280 325 bp mRNA linear EST 13-MAR-1997
DEFINITION HSU46280 Human pancreatic cancer cell line Patu 8988t Homo sapiens
ACCESSION U46280
VERSION U46280.1 GI:1236118
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 325)
Gress,T.M., Muller-Pillasch,F., Geng,M., Zimmerhackl,F., Zehetner
,G., Friess,H., Buchler,M., Adler,G. and Lehrach,H.
TITLE A pancreatic cancer-specific expression profile
JOURNAL Oncogene 13 (8), 1819-1830 (1996)
MEDLINE 97050791
COMMENT Contact: Friederike Mueller-Pillasch
Internal Medicine 1, University of Ulm
Robert-Koch-Strasse 8, Ulm, Baden-Wuerttemberg 89081, Germany
Tel: 0731/502-4311
Fax: 0731/502-4302
Email: friederike.mueller-pillasch@medizin.uni-ulm.de.

FEATURES
source
1..325
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="xs254"
/clone_lib="Human pancreatic cancer cell line Patu 8988t"
/tissue_type="pancreatic cancer"
/cell_line="Patu 8988t"

BASE COUNT 97 a 68 c 73 g 83 t 4 others
ORIGIN

Query Match 13.6%; Score 307.4; DB 10; Length 325;
Best Local Similarity 98.2%; Pred. No. 5e-49;
Matches 319; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 751 ctatcataagactaccacacccaaactgtttacctctggacaatgccacctaccctcaga 810
Db 1 CTATCATAGACCTACCCCAACCACTGTTTACCTCTGGCAATGCCACCTACCTCAGA 60

Qy 811 aactgaaggaggttgattattcaacgcataatggtcggaataatggcaacttgggttttaca 870
Db 61 AACTGAAGGAGGTGGNTATTCAAGCGCATATGGTCGGAATAATGGCACTGGGTTTTTACA 120

Qy 871 gaaagaatgcattcccaacgaagagattgatacctttttt-ggttcccttttggga 929
Db 121 GAAAGAATATGATGCCCAACCAAGAGAGATTGATACCTTTTGGGTTCCTTTTGGGA 180

Qy 930 agtggggattactatcacactacaaaatgtgacagtctctgggatgtgtggtctatgacttg 989
Db 181 AGTGGGATTACTATACACTACAAAATGTGACAGTCTCTGGGATGTGTGGCTATGACTTG 240

Qy 990 tatgaaacgcacaatgctgcctgggaactgatgacaatgggcataatactccacacagatgtac 1049
|||||
```

```
Db 241 TATGAAAACGACAATCTCTNCTGGGATGACAAATGGNATATNCTCCACACAGATGTAC 300
Qy 1050 actcagagagatcacagaaatttag 1074
Db 301 ACTCAGAGAGTACAGGAATCTTAG 325

RESULT 14
LOCUS BM461300 BM461300 1157 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_64119791 NCI_CGAP_Ov44 Mus musculus cDNA clone
IMAGE:5504049 5', mRNA sequence.
ACCESSION BM461300
VERSION BM461300
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1157)
NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Aaron Hsueh
CNA Library Preparation: Life Technologies, Inc.
CNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12144 row: m column: 10
High quality sequence stop: 738.

FEATURES
source
1..1157
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5504049"
/clone_lib="NCI_CGAP_Ov44"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: ovary, PMSG-treated; Vector:
pCMV-SPORT6.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dT. Average insert size
2.2 kb. Library constructed by Life Technologies. Note:
this is a NCI_CGAP Library."

BASE COUNT 351 a 260 c 251 g 294 t 1 others
ORIGIN

Query Match 12.6%; Score 284.6; DB 10; Length 1157;
Best Local Similarity 75.1%; Pred. No. 8.1e-45;
Matches 402; Conservative 0; Mismatches 114; Indels 19; Gaps 3;

Qy 1716 gactgggtcccccctcagctctttcagcaacctggcagcaacgggtggcacatgaacgg 1775
Db 21 GACTGGGTCCCCCTCAGGCTTTC-GCAATCTGGCCCCAACCCGGTGGCACAATGAAGG 79

Qy 1776 atcacctcgtcaactggcaaaagtgtatggttttcaacatcacagcaccatagag 1835
Db 80 ATTACTTGTCAACTGGGAAGAGATATCTGGCTTTTCAACATCACGGCTGATCCGTATGAG 139

Qy 1836 aggggtggacctatcaacaggtatccaggaatcgtgaagaagctctcagggagctcca 1895
Db 140 AGGGTGGACCTATCCAGCAGGTATCCAGCATCGTGAAGAAGCTGCTGCGGAGACTGTCA 199

Qy 1896 cagttcaacaaaatgcagtcgagtcaggtatcccccaagaccccaagagtaacctc 1955
Db 200 CAATTCAACAAGACTGCGGTGCTGTCAGGTACCCCCCGCAGGATCCCAAGAACCAACCT 259

Qy 1956 aggctcaatgaggggtctagggaccatggtatagaggggaaacaaagaaagagcca 2015
Db 260 AGGCTCAACGGAGGAGTCTGGGGTCCGTGTTAAAGAGGAAAAACAAGAAAGAGTCA 319
```



```
Qy 2016 agcaaaatcaggctgagaaagcaaaagaaaaa-----aaagaagaagaaacag 2071
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 320 AACAAACCAAGGCTAAGAAATGCAAAAGAAAGAAAGTAAGGCAAGGATGAGGAAGCAG 379
Qy 2072 cagaagaagctcaggttcaacttgccatcaggtgtaacttggtaagcacaata 2131
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 TTGGCAGCTCACTCCTCAATAAATGCCACCAAGTGTGGCAGCTGGTAAATGCAAAAC 439
Qy 2132 ttctcgtttggttaaaacttaatacagttcttattcttctcagttcttctcaggttaaaacca 2191
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 440 TTTTCGCTTGGTTAAACTTAGA-----CTGCCTGGTTCTTAGGTAATTCA 485
Qy 2192 goaaatttggtgcataatcgcctgagcctgaagcgtcaggctgttttctcagctg 2246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 486 AAGAAATTGGCTCAGAAATACCTCGGCCAACCGCTCGGGCTGTGTTTCACATTG 540

RESULT 15
BE553565 628 bp mRNA linear EST 15-AUG-2000
LOCUS ur47a08.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153398 5'
DEFINITION similar to SW:ARSB_FELCA P33727 ARYLSULFATASE B PRECURSOR ;, mRNA
sequence.
ACCESSION BE553565
VERSION BE553565.1 GI:9817982
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LInL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1056154
Seq primer: -40RP from Gibco
High quality sequence stop: 390.
FEATURES
source
1..628
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 141 a 201 c 165 g 121 t
ORIGIN
```

```
Query Match 11.6%; Score 262; DB 10; Length 628;
Best Local Similarity 64.6%; Pred. No. 2.1e-40;
Matches 407; Conservative 0; Mismatches 220; Indels 3; Gaps 1;
```

```
Qy 596 gggttaccagagatcgagattaaacacctactcttgacagctcgctgcgaaggaggt 655
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GGGATACCATGCGCTCGGATATCGAGACCCCAACCGCTGGACCGGCTAGCAGCTGAGGGGCT 60
Qy 656 taaactggagaaactatgctcagctctatttgcacacatccacgaagtcagttttatcac 715
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CAACCTGGAGAATTATTACATCCAAACCCCATATGTTACACCTTCGAGGAGTCAACTCCTCAC 120
```

```
Qy 716 tggaagtatcatagatacacacccggacttcaacattcttatcataagacatccccacccaa 775
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CGGCAGGTACAGATCCATACAGGATTGCAGCACTCCATTATCCGCCACAGCGCAGCCCAA 180
Qy 776 ctgtttacctctgacaatgcccacctacacctcagaaaaagagggttgatatatcaac 835
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 CTGCTGCCCCGTGGACCAAGGTGACGCTGCCCAAGAGCTACAGGAGGCAAGGCTACTCCAC 240
Qy 836 gcatagtcggaaaaatggcacttgggtttttacagaaaaaagatgcatgcccaccagaag 895
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CCACATGTTGGCAAGTGGCATCTCGGGTTCTACCGGAAGGAGTGTTCCTTACCTACCGCGC 300
Qy 896 aggtattgataccttttttgggttcccttttgggaagtggggaattactatcacacactaca 955
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GGGCTTCGACACACTTCCTGGGTTCCTCAGAGCAATGTGGATTACTACACCTACGACAA 360
Qy 956 atgtacagtcctgggatgtgtgctatgacttgtatgaaaaacgacaatgctgcctggga 1015
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CTGTGATGGCCCAAGGGGTTTGGGTTTGACCTGCACGAGGTGAGAGCGTGGCCTGTG - 419
Qy 1016 ctatgacaatggcatatatactccacacagatgtacactcagagagtacagcaaatcttagc 1075
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 --GGCTCAGCGGCCAGTACTCCACTATGCTCTACGCTCAGCGTGCAGCCACATCTCTGC 477
Qy 1076 ttccataacccccacaagcctatatattttatatatttgcctatcaagctgttcttacc 1135
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 478 CAGGCACAATCCCCAGAAATCCCTCTCTTCTCTATGCTGGCCTTCCAGGCAAGTACACAGCG 537
Qy 1136 actgaaactcctgcaggattattcgaacactacacgataccattatcaacataaacaggag 1195
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 538 CTACAGTCACTTCGAGAAATACCTATACCGCTTACCGCAGATGGGCAACGTTAGCAGACGC 597
Qy 1196 gagatatgctgcatgcttctcctgcttaga 1225
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 598 CAAGTACCGCAGCATGGTGACCTGCATGGA 627
```

Search completed: July 23, 2002, 15:52:09
Job time: 3130 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:08:04 ; Search time 2392.27 Seconds
(without alignments)
12784.552 Million cell updates/sec

Title: US-09-495-823-8
Perfect score: 2266
Sequence: 1 cacggtcgccacgcgc.....tgccactgtgcgaattc 2266

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0 27472414

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estmu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_htc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_htc:*
 - 12: gb_gss:*
 - 13: em_gss_hum:*
 - 14: em_gss_inv:*
 - 15: em_gss_pln:*
 - 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	505	22.3	817	10 BG680752	BG680752 602628445
C 2	465	20.5	615	9 AW474222	AW474222 xs20h05.x
3	382	16.9	683	10 BF307503	BF307503 601893984
C 4	212	9.4	735	10 BG772761	BG772761 602720927
5	210	9.3	344	10 W81485	W81485 zd87a04.r1
6	201	8.9	547	10 BG661302	BG661302 ia59e07.y
C 7	200	8.8	1032	10 BF979627	BF979627 602287919
8	155	6.8	274	10 BG194726	BG194726 RST13892
9	119	5.3	552	10 BF218442	BF218442 601881811
10	115	5.1	325	10 U46280	U46280 HSU46280 Hu
C 11	97	4.3	414	10 W81455	W81455 zd87a04.s1
C 12	94	4.1	458	10 R43179	R43179 yg18a03.s1
C 13	94	4.1	463	9 AA159953	AA159953 zo75a09.s
14	81	3.6	849	10 BF696654	BF696654 602124758
15	70	3.1	407	9 AA159952	AA159952 zo75a09.r
16	55	2.4	951	10 BG167651	BG167651 602342825
17	53	2.3	625	12 AZ815176	AZ815176 2M0083G13

18	53	2.3	648	9 BB520951	BB520951 BB520951
19	53	2.3	700	9 BB319509	BB319509 BB319509
20	40	1.8	265	9 BE149009	BE149009 CM0-HT024
21	33	1.5	61	10 R20122	R20122 yg18a03.r1
22	30	1.3	459	10 BF653682	BF653682 277595 MA
23	29	1.3	388	9 BE047859	BE047859 tz43c06.y
24	28	1.2	717	10 BG975362	BG975362 602843009
25	27	1.2	338	10 B1693994	B1693994 603342563
26	27	1.2	397	10 BG899940	BG899940 HOA40-1-E
27	27	1.2	552	9 AL022653	AL022653 s7827b33
28	27	1.2	575	9 AV596948	AV596948 AV596948
29	27	1.2	625	10 BF537953	BF537953 602049394
30	27	1.2	692	10 B1328079	B1328079 602979220
31	27	1.2	724	10 B1697980	B1697980 603344126
32	26	1.1	185	9 AW429593	AW429593 67921 WAR
33	26	1.1	277	9 AV615802	AV615802 AV615802
34	26	1.1	337	10 BM495947	BM495947 IPCGBT5_2
35	26	1.1	479	9 AI641294	AI641294 fcl3b07.y
36	26	1.1	506	9 AI497032	AI497032 fb59c12.y
37	26	1.1	514	9 AI877708	AI877708 fc50g02.y
38	26	1.1	543	9 AV617615	AV617615 AV617615
C 39	26	1.1	565	10 BM280523	BM280523 ki01h08.y
40	26	1.1	602	9 AV595356	AV595356 AV595356
41	26	1.1	604	9 AV595699	AV595699 AV595699
42	26	1.1	703	10 B1109023	B1109023 602896840
43	26	1.1	774	10 B1689140	B1689140 603315947
44	26	1.1	804	9 AI545901	AI545901 fb76h10.y
45	26	1.1	806	10 B1684551	B1684551 603306926

ALIGNMENTS

RESULT 1

LOCUS BG680752 817 bp mRNA linear EST 01-MAY-2001
DEFINITION 602628445F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4753259 5', mRNA sequence.
ACCESSION BG680752
VERSION BG680752.1 GI:13912149
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 817)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM0612 row: j column: 12
High quality sequence stop: 789.
Location/Qualifiers
1. 817
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4753259"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1. Not I; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5Kb. Library constructed by Life Technologies. Note: this is a NCI_CGAP Library."

FEATURES
source

BASE COUNT 259 a 193 c 207 g 158 t

ORIGIN

Query Match 22.3%; Score 505; DB 10; Length 817;
Best Local Similarity 99.7%; Pred. No. 3.1e-197;
Matches 605; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1368 tgggaaggaggtatcggtgctgtagcttctgtgtagccctctctgaaacaaggga 1427
|||||
Db 9 TGGGAAGGAGGATCGGGCTGTAGCTTGTGTCATAGCCCACTTCTGAAAAACAGGGA 68
|||||

Qy 1428 acagtgtgaaggaacctgtgcacatcactgactggtgaccccaactctctatttcaactggct 1487
|||||
Db 69 ACAGTGTGTGAAGGAACCTGTGCACATCACTGACTGGTACCCCACTCTCATTTCACTGGCT 128
|||||

Qy 1488 gaagacagattgatgagacaattcaactagatggtctgatatctgtggagaccataagt 1547
|||||
Db 129 GAAGGACAGATTGATGAGACATTCAACTAGATGGCTATGATATCTGGAGACCAATAAGT 188
|||||

Qy 1548 gaggtcttcgctcaccgagtagatatatttgcataaactgaccccatatcaccaag 1607
|||||
Db 189 GAGGCTCTCGTCAACCCGAGTAGATATTTTGATATACATTGACCCCATATACACCAAG 248
|||||

Qy 1608 gcaaaaaatggtctctggtgagcagcaggtctatgggtctggaacactgcaatccagtcagcc 1667
|||||
Db 249 GCAAAAAATGGCTCCTGGGACGAGGCTATGGGATCTGGAACACTGCAATCCAGTCAGCC 308
|||||

Qy 1668 atcagagtgcagcactggaattgcttacaggaatactctggtacagagactgggtccccc 1727
|||||
Db 309 ATCAGAGTGCAGACTGGAATTTGTTTACAGGAATCTCTGGCTACAGCGACTGGGTCCCC 368
|||||

Qy 1728 cctcagctcttcagcaacctgggacccgagcagcgtggtgcaaatgaaacggtacccctcgta 1787
|||||
Db 369 CCTCAGTCTTTTCAGCAACTGGGACCGACCGTGGCACAATGAACGGATCACTTGTC 428
|||||

Qy 1788 actgcaaaagtgtatggttttcaacatcacagcagccacatabgaggtggaacct 1847
|||||
Db 429 ACTGCAAAAGTGTATGGCTTTTCAACATCACAGCGACCCCATATGAGAGGTTGGACCTA 488
|||||

Qy 1848 tctaacaggtatccaggaactgtgaagctcctcagaggtctcagaggtctcagagttcaacaaa 1907
|||||
Db 489 TCTAACAGGTATCCAGGAATCTGGAAGAGCTCTTACGAGGAGCTCTCACAGTTTCAACAAA 548
|||||

Qy 1908 actgcagtcggtcaggtatcccccccaagaccccaagaaagtaacccctaggtcgaatgga 1967
|||||
Db 549 ACTGCAGTCCGGTCAGGTATCCCCCAAGACCCCAAGAGTAACCCTAGGCTCAATGGA 608
|||||

Qy 1968 ggggtct 1974
|||||
Db 609 GGGGTCT 615
|||||

RESULT 2
AW474222/c
LOCUS
DEFINITION
similar to SW:ARSB_HUMAN PI5848 ARYL SULFATASE B PRECURSOR ;, mRNA
sequence.

ACCESSION
AW474222
KEYWORDS
EST.
SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 615)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Gibco
High quality sequence stop: 435.
Location/Qualifiers
1. 615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2770233"
/clone_lib="NCI-CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 116 a 151 c 153 g 195 t
ORIGIN

Query Match 20.5%; Score 465; DB 9; Length 615;
Best Local Similarity 99.8%; Pred. No. 1e-180;
Matches 515; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1459 actggtacccccctctcattcactggctgaaggacagattgatgaggacattcaactag 1518
|||||
Db 593 ACTGGTACCCCACTCTCATTTTTCACCTGGCTGAAGGACAGATTGATGAGGACATTCAACTAG 534
|||||

Qy 1519 atgctatgatctctggtgagaccataagtgggtcttcgtccacccccagtagatatt 1578
|||||
Db 533 ATGGGTATGATATCTGGGAGACCAATAAGTGAAGGTCTTCGCTCACCCCGAGTAGATATTT 474
|||||

Qy 1579 tcataaactgaccccatatcaccaagcacaataatgctcctctgggcagcaggtatg 1638
|||||
Db 473 TCATTAACATTGACCCCATATACCAAGGCAAAAATGGCTCTGGGACGAGGCTATG 414
|||||

Qy 1639 ggatctggaacactgccaatccagtcagccatcagagtcagcagcactggaatgcttcacag 1698
|||||
Db 413 GGATCTGGAACACTGCAATCCAGTCAGCCATCAGAGTCAGCACATGGAATTTGCTTACAG 354
|||||

Qy 1699 gaaatcctggtctacagcagctgggtcccccctcaagtcttccagcaacctgggacgaacc 1758
|||||
Db 353 GAAATCCTGGCTACAGCGACTGGGTCCGCCCTCAGTCTTTCAGCAACCTGGGACCCGAAAC 294
|||||

Qy 1759 ggtggcacaatgaacggatcacctcgtcaactggcacaagtgtaagtgatgcttttcaacatca 1818
|||||
Db 293 GGTGGCACAAATGACGGATCACCTTGTCAACTGGCAAAAAGTGTATGGCTTTTCAACATCA 234
|||||

Qy 1819 cagccgaccccatatgagaggtggtgacctatctaacagggtatccaggaatcgtgaaagaagc 1878
|||||
Db 233 CAGCGACCCCATATGAGAGGGTGGACCTTATCTAACAGGTATCCAGGAATCGTGAAGAAGC 174
|||||

Qy 1879 tctacggaggtctctcaagttcaacaacaaactgcagtcgagtcaggtatccccccaaag 1938
|||||
Db 173 TCCTACGGAGGCTCTCACAGTTTCAACAAAACCTGCAGTGGCGGTACAGGTATCCCCCCCAAG 114
|||||

Qy 1939 acccgaagtaacccctaggtcgaatggaggggtct 1974
|||||
Db 113 ACCCCAGAAATACCCCTAGGCTCAATGGAGGGGTCT 78
|||||

RESULT 3
BF307503
LOCUS
DEFINITION
BF307503
mRNA sequence.
ACCESSION
BF307503

BF307503
601893984F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4139565 5',
mRNA sequence.
BF307503

VERSION	BF307503.1	GI:11254633	RESULT	4
KEYWORDS	EST.		BG772761/c	
SOURCE	human.		LOCUS	
ORGANISM	Homo sapiens		DEFINITION	602720927f1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4837638 5', mRNA sequence.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		ACCESSION	BG772761
AUTHORS	1 (bases 1 to 683)		VERSION	BG772761.1
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		KEYWORDS	GI:14083414
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		SOURCE	EST.
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov Plate: LLM1050 row: g column: 22 High quality sequence stop: 679.		ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 735) NIH-MGC http://mgc.nci.nih.gov/.
FEATURES	Location/Qualifiers		REFERENCE	National Institutes of Health, Mammalian Gene Collection (MGC)
source	1. .683 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4139565" /clone_lib="NIH_MGC_17" /tissue_type="rhabdomyosarcoma" /lab_host="DH10B (phage-resistant)" /note="Organ: muscle; Vector: pOTH7; Site_1: EcoRI; Site_2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."		JOURNAL	Unpublished (1999)
BASE COUNT	203 a 167 c 135 g 178 t		COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLM10770 row: f column: 07 High quality sequence stop: 725.
ORIGIN			FEATURES	Location/Qualifiers
Query Match	16.9%; Score 382; DB 10; Length 683;		source	1. .735 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4837638" /clone_lib="NIH_MGC_97" /lab_host="DH10B" /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgaag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTNN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
Best Local Similarity	100.0%; Pred. No. 1.4e-146;		BASE COUNT	169 a 235 c 185 g 146 t
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			ORIGIN	
QY 532 caacttcacacccagcccatctcttcttcacatccttcttgacagctgcgcgaag 591			Query Match	9.4%; Score 212; DB 10; Length 735;
			Best Local Similarity	100.0%; Pred. No. 1.2e-76;
Db 2 CAACCTCCACCTCCCGAGCCCATCTCATTTTCATCCTAGCGGATCATAGGATTAGAG 61			Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
			QY 44 ttttttttcttggtggaagctctctagggaggggggagagagagaaagtgaac 103	
QY 592 atgtgggtaccaggtatctgagataaacaacctactcttgacagctgcgcgaag 651				
			Db 237 TTTTCTTCTTCTTCTGCGAGAGCTCTCTAGGGAGGGGGAGGAGAGAGTGAAG 178	
Db 62 ATGTGGGTATACACGGATCTGAGATTAAACACCTACTCTTGACAAGCTCGCTGCCGAAG 121				
QY 652 gagttaaaactgagaaactactatgccagcctatttgacacattccagaggtcaagtta 711			QY 104 tgtgctggaagagcgagccctcttcttcgcgagtcctccattagccatcac 163	
			Db 177 TGTGCTGAGAGACGAGCCCTCTTGTCTTCCGAGTCCCATTAAGCCATCAC 118	
Db 122 GAGTTAAACTGGAGAACCTACTATGTCTCCAGCCTATTGTGCACCACTCCAGGAGTCAGTTTA 181				
QY 712 ttactggaagtatcagatacacacccgacttcaacattctatcataagaccatcccaac 771			QY 164 ttctggaagattaaagttgtcggacatggtgacagctgagagagagagattcttgc 223	
			Db 117 TTCTGGAAGATTAAAGTTTCTCGGACATGCTGACAGCTGAGAGGAGGAGGATTTCTTGC 58	
Db 182 TTACTGGAAGATATCAGATACACCGGACTTCACATTTCTATCATAGACCTACCCAAAC 241				
QY 772 caaactgtttacctctgacaaatgccacccctacctcagaaactgaagaggttgatatt 831			QY 224 caggtggagagctcttcacgcgtctgtgggtgc 255	
			Db 57 CAGGTGGAGAGTCTTACCCTCTGTGGGTGC 26	
Db 242 CCAACTGTTTACTCTGGACAAATGCCACCTACCTCAGAAAATGAAAGGTTGGATATT 301				
QY 832 caacccatagtcggaatggagacttgggtttttacagaaaagatgcattgcccacca 891				
Db 302 CAAACCATATGTCGGAATGGCAATGGCACTTGGGTGTTTACAGAAAAGATGCATGCCACCA 361				
QY 892 gaagaggatttgataccttttt 913				
Db 362 GAAGAGGATTGATACCTTTT 383				

IMAGE:347598 5' similar to SW:ARSB_HUMAN P15848 ARYLSULFATASE B
PRECURSOR ;, mRNA sequence.
ACCESSION W81485
VERSION W81485.1 GI:1392515
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 344)
REFERENCE Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 546 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 303.
FEATURES Location/Qualifiers
1..344
/organism="Homo sapiens"
/db_xref="GDB:1272973"
/db_xref="taxon:9606"
/clone="IMAGE:347598"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAAGTGGGAGCGGCCATCTTTTCTTTTCTTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHH19W."
BASE COUNT 92 a 93 c 82 g 74 t 3 others
ORIGIN

Query Match 9.3%; Score 210; DB 10; Length 344;
Best Local Similarity 99.4%; Pred. No. 1.le-75;
Matches 310; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1551 ggtcttcgctaccccgagtagattttgcatatacttgaccatatacaccaaggca 1610
|||||
Db 1 GGCTTCGCTCACCCGAGTAGATTATTGTCATACATTGACCCCATATACACCAAGGCA 60

Qy 1611 aaaaatggctcctggcgagcgagctatggatctggaacactgcaatccagtcagccatc 1670
|||||
Db 61 AAAAAATGGCTCTGGGACGAGCGCTATGGGATCTGGAACACTGCAATCCAGTCAGCCATC 120

Qy 1671 agatgcagactgaaattgcttacaggaatactctggtacagcagctgggtcccccct 1730
|||||
Db 121 AGAGTGCAGCACTGGAAATTTGTTACAGGAATTCCTGGCTACAGGACTTGGGTCCCCCT 180

Qy 1731 cagttcttcagaaactgggacccgagctgggcacatgaacgactcacctcgtcaact 1790
|||||
Db 181 CAGTCTTTACGACCTGGGACCCGACCCNGTGGCAATGAACGGATCACCTTGTCAACT 240

Qy 1791 ggcaaaagtgtatgctgttttcaacatcacagcccatatgagaggtggacatctct 1850
|||||
Db 241 GCCAAAGTGATGGCTTTTCAACATCACAGCGCCCATATGAGAGGTGGACCTATCT 300

Qy 1851 aacaggtatcca 1862
|||||
Db 301 AACAGGTATCCA 312

RESULT 6
BG661302
LOCUS ia59e07.y1 Melton Human Islets H1z1 Homo sapiens cDNA 5', mRNA
DEFINITION sequence.
ACCESSION BG661302
VERSION BG661302.1 GI:13816937
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 547)
REFERENCE Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
M., Gibbons M., McCann,R., Cole,R., Tsagataishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
JOURNAL Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
COMMENT Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 482.
FEATURES Location/Qualifiers
source 1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Human Islets H1z1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="TOP10"
/note="Organ: Pancreas; Vector: pZeo-2; Site.1: Not I;
Site.2: Xho I; Library constructed using SuperScript
Plasmid Library kit (Life Technologies). cDNA made by
oligo-dT priming. Xho I site destroyed during cloning.
Size selected by column fractionation; average insert
size 1.59 kb. Primary library, unamplified."
BASE COUNT 169 a 124 c 120 g 134 t
ORIGIN

Query Match 8.9%; Score 201; DB 10; Length 547;
Best Local Similarity 100.0%; Pred. No. 4.5e-72;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2052 aaaaaaagaagaacacagcagagaagcagctctcaggttcaacttgccattcagggttta 2111
|||||
Db 157 AAAAAAAGAGAGAAACAGCAGAGAAGCACTCTCAGGTTCAACTTCCATTCCAGGTGTTA 216

Qy 2112 cttgtggataagcacaataatttcctgttgggttaaacatttaacagttcttatcttta 2171
|||||

Db 136 GTCCATCCATTAAAGCATCATTCTGGAAGATTAAAGTTGTTCGGACATGTTGACAGCTG 77

Qy 202 agagagagaggattctt 221
 |||
 Db 76 AGAGGAGAGGAGGATTCTT 57

RESULT 8
 BG194726
 LOCUS RST13892 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence. EST 21-APR-2001
 DEFINITION
 ACCESSION BG194726
 VERSION BG194726.1 GI:13716413
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 274)
 AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
 Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
 Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
 ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
 ,J., Danzig,J. and Ducar,M.
 TITLE Creation of genome-wide protein expression libraries using random
 activation of gene expression
 JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
 MEDLINE 21227151
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@atersys.com
 High quality sequence stop: 224.
 Location/Qualifiers
 source
 1..274
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."
 BASE COUNT 77 a 65 c 55 g 76 t 1 others
 ORIGIN

Query Match 6.88; Score 155; DB 10; Length 274;
 Best Local Similarity 100.0%; Pred. No. 5e-53;
 Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 790 acatgcacacctaccatcagaactgaaggaggttgatattcaacgcataatggtcgaa 849
 |||
 Db 120 ACATGCCACCTACCTCAGAACTGAAGGAGGTGGATATTCAACGCATATGTCGGAA 179
 |||

Qy 850 aatggcacttgggtttttacagaaagaatgcgtccaccagagagattgataacct 909
 |||
 Db 180 AATGGCACTTGGGTTTTACAGAAAGAATGCATGCTCCACCAGAGAGATTGTATACCT 239
 |||

Qy 910 ttttggttcccttttgggaagtgggattactat 944
 |||
 Db 240 TTTTGGTTCCCTTTTGGGAAAGTGGGATTACTAT 274
 |||

RESULT 9
 BF218442
 LOCUS BF218442 552 bp mRNA linear EST 06-NOV-2000
 DEFINITION 601881811F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4094211 5',
 mRNA sequence.


```

/db_xref="taxon:9606"
/clone="IMAGE:347598"
/clone_lib="Soares_fetal_heart_NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHH19W."
BASE COUNT      74 a 105 c 105 g 130 t
ORIGIN

Query Match      4.3%; Score 97; DB 10; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.1e-29;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1625 ggcagcggctatggatctggaaactgcacatccagtcagatgcagcactg 1684
|||||
Db 413 GGCAGCGGTATGGATCTGGACACTGCATCCAGTCAGCCATCAGATGCAGCACTG 354
|||||

QY 1685 gaaattgcttacgagaatctcgtgctacagcagctgg 1721
|||||
Db 353 GAAATTGCTTACAGGAATCTCGGTCTACAGCGACTGG 317
|||||

RESULT 12
R43179/c
LOCUS      R43179      458 bp      mRNA      linear      EST 22-MAY-1995
DEFINITION yq18d03.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:32854 3', similar to SP:ARSB_HUMAN P15848 ARYLSULFATASE B
PRECURSOR ;, mRNA sequence.
ACCESSION  R43179.1 GI:825405
VERSION     EST.
KEYWORDS    human.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 458)
AUTHORS     Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohldmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
COMMENT     On May 8, 1995 this sequence version replaced gi:801403.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1275
High quality sequence stops: 384 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1275 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 384.
FEATURES
source      Location/Qualifiers
1..458
/organism="Homo sapiens"
/db_xref="GDB:405201"

```

```

/db_xref="taxon:9606"
/clone="IMAGE:32854"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: whole brain; Vector: Lfamid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AAGTGAAGAATTCGCGCGCGCGCAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      74 a 110 c 119 g 152 t
ORIGIN

Query Match      4.1%; Score 94; DB 10; Length 458;
Best Local Similarity 100.0%; Pred. No. 5.2e-28;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1881 ctacggggctctcacagttcaacaaactgcagtcgcgtcaggtatcccccaagac 1940
|||||
Db 179 CTACGGAGGCTCTCACAGTTCAACAAACTGCAGTGGCGGTATCCCCCAAGAC 120
|||||

QY 1941 ccagaagtaaccctagctcaatggagggtct 1974
|||||
Db 119 CCCAGAAGTACCCTAGGCTCAATGGAGGGGTCT 86
|||||

RESULT 13
AA159953/c
LOCUS      AA159953      463 bp      mRNA      linear      EST 16-DEC-1996
DEFINITION zof5a09.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
IMAGE:592696 3', similar to SW:ARSB_HUMAN P15848 ARYLSULFATASE B
PRECURSOR ;, mRNA sequence.
ACCESSION  AA159953
VERSION     AA159953.1 GI:1734444
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 463)
AUTHORS     Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 369.
FEATURES
source      Location/Qualifiers
1..463
/organism="Homo sapiens"
/db_xref="GDB:4624082"
/clone="IMAGE:592696"
/lab_host="Stratagene pancreas (#937208)"

```

/note="Organ: pancreas; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pancreatic adenocarcinoma cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'".

BASE COUNT 82 a 115 c 115 g 151 t

ORIGIN

Query Match 4.18; Score 94; DB 9; Length 463;
Best Local Similarity 100.0%; Pred. No. 5.1e-28;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1881 ctacgaggtctcacagtccaacaaactcagtcggtcggtcaggtatcccccaagaac 1940

Db 172 CTACGGAGGCTCTCACAGTTCAACAAACTCGAGTCCGGTCAGGTATCCCCCAAGAC 113

Qy 1941 cccgaagaatcccttaggtcaatggaggggtct 1974

Db 112 CCCAGAAGTAACCTAGGCTCAATGGAGGGTCT 79

RESULT 14
BF696654
LOCUS 602124759f1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4281567 5',
DEFINITION mRNA sequence.
ACCESSION BF696654
VERSION BF696654.1 GI:11982062
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 849)

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
Plate: LICM110 row: d column: 16
High quality sequence stop: 579.

FEATURES
Location/Qualifiers

1..849
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4281567"
/clone_lib="NIH_MGC_56"
/tissue_type="primitive neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctcgccc); Site_2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCGACATG-dt(30)BN-3' (where B = A, C, G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 237 a 175 c 180 g 257 t

ORIGIN

Query Match 3.6%; Score 81; DB 10; Length 849;
Best Local Similarity 100.0%; Pred. No. 9.1e-23;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2060 aagaagaacacagcagaagcagtcaggttcaacttcaggttcaggtgtgga 2119

Db 1 AAGAAGAACAGCAGAGCAGTCTCAGGTTCAACTTCCCATTCAGGTGTGGA 60

Qy 2120 taagcacaaatatttcctgtt 2140

Db 61 TAAGCACAATATTCCTGTT 81

RESULT 15
AA159952
LOCUS 2075A09.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
DEFINITION IMAGE:592696 5', mRNA sequence.
ACCESSION AA159952
VERSION AA159952.1 GI:1734443
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 407)

REFERENCE Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chissole, S., Dietrich, N., DuBuque, T., Lacy, M., Le, N., Mardis, E., Moore, M., Hultman, M., Kucaba, T., Prange, C., Rifkin, L., Rohlfing, T., Scheilenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M. Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 274.

FEATURES
Location/Qualifiers

1..407
/organism="Homo sapiens"
/db_xref="GDB:4624082"
/db_xref="taxon:9606"
/clone="IMAGE:592696"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: pancreas; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pancreatic adenocarcinoma cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 82 a 99 c 114 g 104 t

ORIGIN

Query Match 3.1%; Score 70; DB 9; Length 407;
Best Local Similarity 100.0%; Pred. No. 4.1e-18;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 gagagaggagaaagtgaatgtctgtggaagagcgagccctcttcttccggag 142

Db 215 GAGGAGGAGGAGAAAGTGAATGTCTGGAGAGCGAGCCCTCTTGTCTCCGGAG 274

Qy 143 tcccatccat 152

ORIGIN

Db 275 TCCCATCCAT 284

Search completed: July 23, 2002, 17:46:44
Job time: 9520 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 17:46:49 ; Search time 133.2 Seconds
(without alignments)
714.318 Million cell updates/sec

Title: US-09-495-823-7
Perfect score: 3012
Sequence: 1 MAPRCAGHPPPPSPQACV.....VPRYPKPRSPNRLNGV 550

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhcc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	SUMMARIES				Description
	Score	Query Match	Length	ID	
1	1293.5	42.9	503	5 Q9NJU8	Q9nj8 helix pomat
2	920	30.5	542	5 Q9V6F2	Q9v6f2 drosophila
3	846.5	28.1	579	5 Q9VVM4	Q9vvm4 drosophila
4	752	25.0	996	5 Q9VVK9	Q9vvk9 drosophila
5	646.5	21.5	266	5 Q9NJU7	Q9nju7 helix pomat
6	556	18.5	486	5 Q9VVM1	Q9vvm1 drosophila
7	507	16.8	169	4 Q9H5D8	Q9h5d8 homo sapien
8	472	15.7	521	16 Q9A922	Q9a922 caulobacter
9	447	14.8	520	11 Q9JHK9	Q9jkh9 mus musculus
10	443	14.7	520	11 Q9JKU8	Q9jku8 mus musculus
11	432	14.3	554	3 Q9C0V7	Q9c0v7 schizosacch
12	395.5	13.1	537	16 Q92UC0	Q92uc0 rhizobium m
13	393	13.0	555	16 Q92YF6	Q92yf6 rhizobium m
14	393	13.0	577	2 Q9X759	Q9x759 klebsiella
15	391.5	13.0	525	11 Q9D3B4	Q9d3b4 mus musculus
16	391	13.0	577	2 Q9LAY4	Q9lay4 klebsiella

17	391	13.0	577	2 Q9LAY2	Q9lay2 klebsiella
18	384.5	12.8	525	4 Q96EG1	Q96eg1 homo sapien
19	383	12.7	525	4 Q9Y2K4	Q9y2k4 homo sapien
20	378	12.5	509	4 Q96CJ0	Q96cj0 homo sapien
21	377	12.5	557	16 Q98BQ3	Q98bq3 rhizobium l
22	371	12.3	506	11 Q9DC66	Q9dc66 mus musculus
23	360	12.0	475	2 Q9F821	Q9f821 escherichia
24	359	11.9	559	5 Q16138	Q16138 heliocidari
25	355.5	11.8	551	5 Q25384	Q25384 hemcentrot
26	344.5	11.4	303	5 Q18337	Q18337 caenorhabdi
27	336	11.2	545	16 Q92ZH5	Q92zh5 rhizobium m
28	321.5	10.7	517	2 Q9LSW0	Q9lsw0 prevotella
29	319.5	10.6	787	16 P95059	P95059 mycobacteri
30	285	9.5	512	2 Q93P98	Q93p98 microscilla
31	269	8.9	453	16 Q9CKE0	Q9cke0 pasteurella
32	256	8.5	452	5 Q18924	Q18924 caenorhabdi
33	251	8.3	649	10 Q9ATG5	Q9atg5 chlamydomon
34	248	8.2	382	4 Q9UHJ8	Q9uhj8 homo sapien
35	243.5	8.1	970	16 Q65931	Q65931 mycobacteri
36	236	7.8	524	5 Q9VE24	Q9ve24 drosophila
37	230	7.6	249	2 Q93P96	Q93p96 microscilla
38	221.5	7.4	503	16 Q9I799	Q9i799 pseudomonas
39	219.5	7.3	709	5 Q21376	Q21376 caenorhabdi
40	218	7.3	492	5 Q95R73	Q95r73 drosophila
41	218	7.2	430	2 Q9KYY2	Q9kyy2 streptomyce
42	213.5	7.1	787	16 Q06776	Q06776 mycobacteri
43	212	7.0	538	16 Q91IE5	Q91ie5 pseudomonas
44	207	6.9	202	2 Q93P97	Q93p97 microscilla
45	206.5	6.9	867	13 Q90XB6	Q90xb6 coturnix co

ALIGNMENTS

RESULT 1

Q9NJU8	PRELIMINARY;	PRT;	503 AA.
ID	Q9NJU8		
AC	Q9NJU8;		
DT	01-OCT-2000 (Tremblrel. 15, Created)		
DT	01-OCT-2000 (Tremblrel. 15, Last sequence update)		
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)		
DE	SULFATASE 1.		
GN	SULF1.		
OS	Helix pomatia (Roman snail) (Edible snail).		
OC	Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;		
OC	Helicacea; Helicidae; Helix.		
OX	NCBI_TaxID=6536;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20233298; PubMed=10772344;		
RA	Wittstock U., Fischer M., Svendsen I., Halkier B.A.;		
RT	"Cloning and characterization of two cDNAs encoding sulfatases in the		
RL	Roman snail, Helix pomatia."		
RL	IUBMB Life 49:71-76(2000).		
DR	EMBL; AF109924; AAF30402.1; -.		
DR	HSSP; P15848; IFSU.		
DR	InterPro: IPR000917; Sulfatase.		
DR	Pfam: PF00884; Sulfatase; 1.		
DR	PROSITE; PS00149; SULFATASE_2; 1.		
SQ	SEQUENCE 503 AA; 55821 MW; AC45334BA694413D CRC64;		

Query Match 42.9%; Score 1293.5; DB 5; Length 503;
Best Local Similarity 48.3%; Pred. No. 8.7e-102;
Matches 25; Conservative 82; Mismatches 141; Indels 51; Gaps 8;

Qy	37	CLLFTYGLSMQALLEEBCGALQAQAKLEPSTTSQPHLIFILADDQGRDVGVHGS	96
Db	4	CLLVLIATACAVADSS----ASACTROD-----AGQPNIVFLADDFGHDVGVHGS	54
Qy	97	EIKPTPTLDKLAAGVKNLENYVQICTPSPRSQFTGYQIHTGLQHSIRTPQNCPLD	156
Db	55	EIHTPTLDALSASGVRLNENYVQICTPSTRSOLMSGRYQIHTGLQHSIRTPQNCPLD	114

```
Qy 157 NATLPQKLKEYGYTHMVGKHLGFYRKECMPTRRGDTFFGSLGSGDYTH----- 209
: || |||| ||-||| ||||| ||||| : ||||| : || : || :
Db 115 SPTLADKLKESGYATHMVGKHLGFYKOEYLPWNRGDTFFGYLNAEDYFNHNVPWROV 174
: || |||| ||-||| ||||| ||||| : ||||| : || : || :
Qy 210 -YKDSFGMCGYDLYENDNAWDYNGIYSTQMTQVQQLASHNPTKPIFLYIAQAV 268
: || |||| ||-||| ||||| ||||| : ||||| : || : || :
Db 175 RY-----LDRDN-NGPVRNETGOYSAHLFTGRAIDVQVSHNTSKPLFLYLAQSV 224
: || |||| ||-||| ||||| ||||| : ||||| : || : || :
Qy 269 HSPLOACRPFYEHYRSININRRYAAALSCLDIAINNVTALATYGFYNNIIYSSDN 328
: |||| : || : |||| : |||| : |||| : |||| : |||| : |||| :
Db 225 HAPLEVPKEYEHKRYNTDKNRRTFAGVYSALDEGVANLTQALKDGLNNVTILFSTDN 284
: |||| : || : |||| : |||| : |||| : |||| : |||| : |||| :
Qy 329 GQPTAGSNMPLRGSGTYGWEIGRAVGFVHSPLLKNGKTVCKPEVHITDWTPLISLA 388
||| |||| : |||| : |||| |||| |||| : |||| : |||| : |||| :
Db 285 GQIHAGNNYPLRGKASLWEGFVGFGVSGALRSGAVSKGLTHVSDWFTPLVILA 344
||| |||| : |||| : |||| |||| |||| : |||| : |||| : |||| :
Qy 389 EQIDEIDLDGYDIWETISSEGLRSPRVLDLHNTD-----PIYTKANGNSWAAGYGI 440
: |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Db 345 CGNLNGTKPLDGFNQWDITISNETSPREILLHNTDILYPOKGVPLYSNT----- 393
: |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Qy 441 WNTAIQSAIRVQHVKLLTNGPGYSDWPPQSFNSLGNRWHNERITSTGKSVHFNITA 500
: |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Db 394 WDTRVRAAIRVGDYKLLTGDPGNSWPPPPD-----GHLYEVPFIEQSAKNWLEFNITA 448
: |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Qy 501 DPYERVDLSNRYPGVIAKLLRLSQFNKTAVPVRYPPKDRSPNRLNGV 550
: |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Db 449 DPNEHNDLSSEKPLEVLRLQLLQVFNNTAVPPRYPPAPDPCDPAALHGDV 498
: |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
RESULT 2
Q9V6F2 PRELIMINARY; PRT; 542 AA.
AC Q9V6F2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG8646 PROTEIN.
GN CG8646.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN 1;
RP SEQUENCE FROM N.A.
RC SPRAIN-BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencs P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry J., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

```
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003821; AAF58475.1; -.
DR HSP; P15848; IFSU.
DR FlyBase; FBgn0033763; CG8646.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
SQ SEQUENCE 542 AA; 60278 MW; EACD9C14DD2B3204 CRC64;

Query Match 30.5%; Score 920; DB 5; Length 542;
Best Local Similarity 38.1%; Pred. No. 7.3e-70;
Matches 213; Conservative 79; Mismatches 169; Indels 98; Gaps 15;

Qy 34 WILCLLYGYLSWGQALKEEBEGALLAQAGEKLEPSTSTSQPHLIFLADDDQGFROVGY 93
: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 5 WLICLL-----LPIIDAAEVEK-----SPAKNIIIFLADDLGFDVGF 43
: ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Qy 94 HGS-EIKPTLTKLAAEGVKLENYVQICPTPSQFTGKYQIHTGLQHSIIIRPTQNC 152
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 HGSAEITPNIDALAYSGIILNRYVAPICPTPSALMTGKYPIHTGQHVLYAAERPG 103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 153 LPLDNATLPQKLKEYGYTHMVGKHLGFYRKECMPTRRGDTFFGSLGSGDYTHYKC 212
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 LPLEEKILPQYLNELGYTSHIAGKHLGHWLKVTPLYRGFSSHGLDMRNGT----- 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 213 DSPCMCGYDLYENDNAWDYNGIYSTQMTQVQQLASHNPTK-PIFLYIAQAVHS- 270
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 ----QVAYDLH-----GHYTTDVTIDHSVKVIANHNATKGPLFLYVAHAACHSS 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 271 ----PQAQGRYFEHYRSIININRRYAAALSCLEAINNVTALATYGFYNNIIYSS 326
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 NPYNLPVPDNDVIKMSHIPNKKRKAAMYSKMDNSGVQIVDQLRKSNNLENSIIIFSS 261
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 327 DNGQOPTAG-----GSNNPLRGSGTYGWEIGRAVGFVHSPLLKNGKTVCKPEVHITD 381
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 DNGG-PAQGFNLNPNFASNYPLKGVKNTLWEGVRAAGLWSPLLKKKSQVSNQTMHIIDL 320
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 382 PTLISLAEGQ---IDEDIQLDGYDIWETISSEGLRSPRVLDLHNTDPIYTKAK----- 430
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 PTLLEAAGGQPALNLSKQIDQSIWRAVQDKASPRUNLVJHNTDIDWGSAAALSVDGDKL 380
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 431 -----NGSAAAGYGIWNTAIQSAIRVQHVKLLTNGPGYSDWPPQSFNSLGNRWHNER 484
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 VKGTNYRGSNDGWG---PAGEROPRLYDQVQV---GRSRAGRALEALKMLPSRADQOR 433
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 485 I-----TSSTGKSV-----WLFNITADPYERVDLSNRYPGIIVKLLRLRLRSQ 526
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 IRAAATVSCPGQSGQTSCTAFAPCLFIHRDDPCQYLNLAQYEPVNNALMTLELRF 493
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 527 NKTAVPVRYPPKDRSNR 545
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 494 NATAVPPSNKPADPRADPR 512
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
Q9VYM4 PRELIMINARY; PRT; 579 AA.
ID Q9VYM4
AC Q9VYM4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
```



```
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000);
DR EMBL; AE003522; AF49290.1; -.
DR HSSP; P15848; IFSU.
DR FlyBase; FBgn0036765; CG7408.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
SQ SEQUENCE 486 AA; 55063 MW; 4327F6726A98F524 CRC64;
```

```
Query Match      18.5%; Score 556; DB 5; Length 486;
Best Local Similarity 31.3%; Pred. No. 7.1e-39;
Matches 148; Conservative 65; Mismatches 166; Indels 94; Gaps 13;
```

```
QY 140 LQHSIIPTOPNCLPIDNATLPQKLKEVGYSTHVMGKWHLGFYRKECMPTRGFDTFEGS 199
Db 1 MQHVIVNDQFQVLPNETTMAEIRFENGXYRTSLLGKWHLGSQRNFTPTPERGDFRHLG 60

QY 200 LLGSGDYTH-YKCDSPCMGCDLYENDNAAWDYDNGIYSTOMYTORVQOITLASH---NP 255
Db 1 LGAVVDYTSYEQQNGYNGHDFRDSLKSTH-DH-VGHYVTDLLTDAVKEIDHGSKNS 119

QY 256 TKPIFLYIAYOAVHS-----PLQAPGRYFEHYRSIINRRRYAAMLSCLEAINNVTLA 310
Db 120 SQPLFLLLNLHAPHAANDDDPMQAPAEVSRFEYSKTHRYAAMVSRLDKSVGSVIDA 179

QY 311 LKTYGFYNNIIYSSDNGQPTAG-----GSNWPLRGSKGTWEGGIRAVGFVHSPLLK 365
Db 180 LARQEMIQNSIIILFSLDNGG-PTQGOHSTTASNPLRGQKSPWEGALRSAAITWFEFE 238

QY 366 NKGTVCKPEPHITDWTYPTLISLABQDDEIDQLDGYDIWETISGLRSPRVDIILHNDP- 424
Db 239 RLGSVWQKQIYIGDLLPTLAAAGISDPDPAHLHDLGLNLSALKYGYSEVEIHHVVIDE 298

QY 425 -----TYTKAKNGSWAAGYGIWNTAIQSAIRVOHQWKLTTG--NPG-YSDWVPPQSF 475
Db 299 VAEPHLSYTRGK-----WKVISQTTNOGLDGLWGHRETSEV 335

QY 476 GPNRWHNERITSSGKSVW----- 494
Db 336 DPRAVEEELVRNT--SWMLQLOQVSGERNISLRDQSRIECDPDATGKPKPLFLEGPC 393

QY 495 LFNITADPYERVDSLNRYPG--IVKKLLRLRSQFNKTAVPVRYPPKDRSNR 545
Db 394 LFDIEADPCERSNLYAEYQNSTIFLDLWSRQQPAKQAHPPNKNKPGPNCDPR 446
```

```
RESULT 7
Q9H5D8 PRELIMINARY; PRT; 169 AA.
ID Q9H5D8
AC Q9H5D8;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE CDNA: FLJ23548 FIS. CLONE LMG08487.
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027201; BABI5689.1; -.
DR HSSP; P15848; IFSU.
SQ SEQUENCE 169 AA; 18905 MW; 4583C92FFFC3380 CRC64;
```

```
Query Match      16.8%; Score 507; DB 4; Length 169;
Best Local Similarity 96.9%; Pred. No. 2.2e-35;
Matches 93; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 455 KLTGNGPYSDWVPPQSFNSLGNPNRWHNRITTSSTGKSVLFNITADPYERVDSLNRYPG 514
Db 25 RLLTGNFGYSDWVPPQSFNSLGNPNRWHNRITLSTGKSVLFNITADPYERVDSLNRYPG 84

QY 515 IVKKLLRLRSQFNKTAVPVRYPPKDRSNRLNGGV 550
Db 85 IVKKLLRLRSQFNKTAVPVRYPPKDRSNRLNGGV 120
```

```
RESULT 8
Q9A922 PRELIMINARY; PRT; 521 AA.
ID Q9A922
AC Q9A922;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE SULFATASE FAMILY PROTEIN.
GN CC1172.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005795; AAK23156.1; -.
DR HSSP; P15848; IFSU.
DR TIGR; CC1172; -.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
```



```

DR Pfam: PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Signal; Hydrolase.
FT SIGNAL 1 20 POTENTIAL.
SQ SEQUENCE 577 AA; 64155 MW; 7E897EDB2CABD18C CRC64;

Query Match
Best Local Similarity 13.0%; Score 393; DB 2; Length 577;
Matches 130; Conservative 81; Mismatches 188; Indels 162; Gaps 20;

Qy 75 QPHLIFILADQGFQVYHSGSEIKTPTLDKLAEGVKLENYYVQPICTPSRSQFITCK- 133
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
25 RPNVIVIAADMGYSIDSPFGCEIPTNLQAMAEQGRMSQYITSPMSAPARSMLLTGSN 84
Qy 134 -----YQHTGLQHSIIIRPTQPCNCLPLDNATLPOKLKEGVGTHVMGKWHLGFY 182
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
85 NQAGMGMMWYDTIGKEGYELRDT-----RVTTWAERFKDAGYNTLMAGKWHLGTV 138
Qy 183 RKECHPTRRGDTFTFGSLGSGDYTHYKCDSPGMC-----YDLYENDNAAWDYDNGIY 237
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
139 -PGATPKRERGFENHAF-AFMGGCT--SHENDAIP--LGTVEAFHTYTRDGERVSLPDDFY 192
Qy 238 STQMYTQVQOILASHNPTKPIFYIAQVHSPLOAP-----GRYFEHYRSIININ 289
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
193 SSEAYARQMSWIKATPKQPVFAWLAFTAPHDPLQAPDEWIKRFGQYQGYAEVYRQR 252
Qy 290 RRR-----YAAMLSCLDEALNNVTLA 310
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
253 IARLKALGIIHDDTPLPLELDKWEALTPQOQKYTAKVMQVYAAAMIANMDAQIGTLMET 312
Qy 311 LKTYGFYNNSTIIYSSDNGGQTAG-----GSMNPLRGSGKGTWEGGIRAVG 357
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
313 LKQTRKNTLLVLTONGANPAQGFYESTPEFWKQFDNSYDNGVRKGSFVSGPWHAN 372
Qy 358 FVHSPLLKNGKTVCKEPIVHTWPTLISLAGQIDEDIDQD-----Y 401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
373 VSNAP-----YANYHKT--TSAQGGINTDFMISGPIRTHGKIDASTMAVY 416
Qy 402 DIWETISEGLRSPRVLDLHNDPIYTKAKNGSWAAGYGNWNTAOSAIRVQHWKLLTGNP 461
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
417 DVAPTYL-----FAGIDPNKSLAK-----PVLPMIGVSFKRYLTGEV 455
Qy 462 GYSWVPPQSFNSLG-----PNRWHNERITS-----STGKSVW-LFNITADPYERVD 507
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
456 ---QEPR--GNYGVELHQAQVWGVGKMLRLVPRGLTAGDAPWQLFNLHDDPLETHD 509
Qy 508 LSNRYPGIVKKLLRLSOFNK 528
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
510 VAAEHPDRVKAMSEAYEAFK 530

RESULT 15
Q9D3B4 PRELIMINARY; PRT; 525 AA.
AC Q9D3B4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 6330406P08RIK PROTEIN.
GN 6330406P08RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=MEDULLA OBLONGATA;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

```

```

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:695-690(2001).
DR EMBL; AK018132; BAB31086.1; -.
DR HGSP; P15289; IAKU.
DR MGS; MGI:1921258; 6330406P08RIK.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
SQ SEQUENCE 525 AA; 57449 MW; DF889E39866D1E1F CRC64;

Query Match
Best Local Similarity 13.0%; Score 391.5; DB 11; Length 525;
Matches 138; Conservative 89; Mismatches 168; Indels 121; Gaps 26;

Qy 71 TSTSOPIHIFILADQGFQVYHSGSEIK-TPTLDKLAEGVKLENYY-VQPICTPSRSQ 128
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
31 TRAPQNPVIVIAADMGWDLGANNAETKDTNLDKMASEGMRVDFHAAASTCSPSRAS 90
Qy 129 FITQKYQHTGLQHSIIIRPTQPCNCLPLDNATLPOKLKEGVGTHVMGKWHLGFYRKECMP 188
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
91 LTTGLRLGRLNGVTHN-FAVTSVGGLPVNETTLLAEVLGEGYVVTAMIGKWHLG-HHGSYHP 148
Qy 189 TRRGFDTPFG-----SLLGSGDY--YTHYKCD-----SPGMCY-----DLYENDN 227
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
149 NFRGFDYFEGIPYSNDMGCTDAPGYNPPCPACPOQDGLWRNPGRCYTDVALPLYENLN 208
Qy 228 AAWDYDNGIYSTQMYTQR-VQOILASHNPTKPIFYIAQVHSPLO-APGRYFEHYRSI 285
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 IVEQPVNLSGLAQKYAERAVEFIEQASTSGRPFLLYVGOAHMHVPLSVTPPLAHPORQSL 268
Qy 286 ININRRRYAAMLSCLEADAINNVTLALTKTYGFYNNSTIIYSSDNGGQPTA-----GGSNWP 340
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
269 YRASLUREMDSLVGQTKDKVDHVA-----RENTLLMFTGDNG--PWAQKCELAGSVGP 318
Qy 341 LRG-----SKGTWEGGIRAVGFVHSPLLKNGKTVCKEPIVHT-----DWYP 382
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
319 FFGLWQTHOGGSPKQTTWEGGHRVPALAYWP-----GRV---PANVTSTALLSLDIFP 370
Qy 383 TLISLAEGQIDEDIDQDYDIWETISEGLRSPRVLDLHNDPIYTKAKNG-----SW 434
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
371 TVIALAGASLPENRKFDRDYSEV-----LFGKSMQGHVRLFPNSG 412
Qy 435 AAG-YGINWTAIOSAIRVQHWK--LLTGNPGYSWVPPQSFNSLGNPNRWHNERITSSTGCK 491
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
413 AAGEYG-----ALQT-VRLNHYKAFYITGGAKAC-----GSVGPQHHVAPL----- 454
Qy 492 SVWLFNITADPYERVDL---SNRYPGIVKKLLRLS 524
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
455 ---IFNLEDADEGMPLQKGSPEYQEVQLQVYTRALA 487

Search completed: July 23, 2002, 19:06:40
Job time: 4791 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 17:07:14 ; Search time 79.69 Seconds
(without alignments)
766.603 Million cell updates/sec

Title: US-09-495-823-7
Perfect score: 3012
Sequence: 1 MAPRCAGHPPPPSPQACVC.....VPRYPKDRSNRLNGGV 550

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2996	99.5	599	22 AAB85483	Human 23553 sulfat
2	2307.5	76.6	515	20 AAV41701	Human PRO708 prote
3	2307.5	76.6	515	21 AAB44257	Human PRO708 (ONQ3
4	2307.5	76.6	515	22 AAU29061	Human PRO polypept
5	1859.5	61.7	569	22 AAB85482	Human 25278 sulfat
6	1562	51.9	289	22 AAU14421	Human novel protei
7	1415.5	47.0	533	22 AAB51184	Human sulfatase pr
8	1079	35.8	366	22 AAU14185	Human novel protei
9	989	32.8	187	22 AAU25415	Human protein sequ
10	920	30.5	542	22 AAB63962	Drosophila melanog
11	846.5	28.1	579	22 ABB63246	Drosophila melanog

12	752	25.0	996	22 ABB61654	Drosophila melanog
13	603.5	20.0	128	22 AAU23335	Novel human enzyme
14	556	18.5	486	22 AAB63259	Drosophila melanog
15	457.5	15.2	492	22 ABB52501	Escherichia coli p
16	443	14.7	520	22 AAB85485	Pfam consensus seq
17	429	14.2	552	22 AAB85484	Pfam consensus seq
18	424.5	14.1	590	22 AAU78688	Human protein seq
19	422.5	14.0	522	22 AAU41939	Human polypeptide
20	413	13.7	583	22 AAV39920	Human steroid sulp
21	404	13.4	583	22 AAB51185	Human sulfatase pr
22	402	13.3	646	22 AAU41129	Human polypeptide
23	399	13.2	631	22 ABG22561	Novel human diagno
24	394.5	13.1	589	22 AAU39343	Human polypeptide
25	384.5	12.8	525	22 AAB85480	Human 22438 sulfat
26	384.5	12.8	552	22 ABB12445	Human bone marrow
27	378.5	12.6	525	22 AAU12199	Human PRO1777 poly
28	356	11.8	507	22 AAB51183	Human sulfatase pr
29	353.5	11.7	736	22 ABG22560	Novel human diagno
30	336.5	11.2	551	22 AAB51186	Sea urchin sulfata
31	324	10.8	571	22 AAG98936	E. coli growth and
32	319.5	10.6	787	22 AAG81127	Mycobacterium tube
33	255	8.5	1120	22 ABG07425	Novel human diagno
34	248.5	8.3	552	22 AAU39414	Human polypeptide
35	243.5	8.1	602	22 AAU41200	Human polypeptide
36	236	7.8	524	22 ABB69201	Drosophila melanog
37	236	7.8	625	22 ABG20519	Novel human diagno
38	221	7.3	510	22 AAB51182	Human sulfatase pr
39	219.5	7.3	709	22 AAE00435	C. elegans sulfata
40	216	7.2	1689	22 ABG16013	Novel human diagno
41	216	7.2	2417	22 ABG28727	Novel human diagno
42	214.5	7.1	822	22 ABG28571	Novel human diagno
43	214	7.1	502	20 AAU42790	Human sulphamidase
44	214	7.1	502	20 AAU90233	Human sulphamidase
45	214	7.1	502	22 AAE08548	Human sulphamidase

ALIGNMENTS

RESULT 1

AA85483	
ID	AA85483 standard; Protein; 599 AA.
AC	AA85483;
DT	25-SEP-2001 (first entry)
XX	Human 23553 sulfatase polypeptide.
DE	Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;
KW	antiinflammatory; vasototropic; antitumor; gene therapy; human.
OS	Homo sapiens.
XX	
PN	WO200155411-A2.
XX	
PD	02-AUG-2001.
XX	
PF	31-JAN-2001; 2001WO-US03266*
XX	
PR	31-JAN-2000; 2000US-0495823.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
XX	
DR	WPI; 2001-476214/51.
DR	N-PSDB; AAH46863, AAH46867.
XX	
PT	Novel human sulfatase polypeptides useful for treating and diagnosing
PT	sulfatase-related disorders such as cerebrovascular diseases, acute
PT	meningitis, multiple sclerosis, degenerative diseases and tumor
XX	

Claim 9; Fig 15; 180pp; English.

The invention provides 22438, 23553, 25278 or 26212 human sulfatase polypeptides and polynucleotides. The sulfatase genes and polypeptides are useful for treating disorders involving the brain such as cerebrovascular diseases, infections such as acute meningitis, demyelinating diseases including multiple sclerosis, degenerative diseases affecting the cerebral cortex including Alzheimer's disease and Pick disease, spinocerebellar degenerations including spinocerebellar ataxias including Friedreich ataxia, and ataxia telangiectasia, degenerative diseases affecting motor neurons including amyotrophic lateral sclerosis, inborn errors of metabolism such as leukodystrophies, toxic and acquired metabolic diseases, including vitamin deficiencies, and neurocutaneous syndromes (phakomatoses) including neurofibromatosis. The present sequence represents a human 23553 sulfatase polypeptide.

Sequence 599 AA:

Query Match	99.5%	Score	2996	DB	22	Length	599
Best Local Similarity	99.6%	Pred. No.	5.3e-276				
Matches	548	Conservative	0	Mismatches	2	Indels	0
QY	1	MARPGCAGHPPSPQACVCPGKMLANGALAGFWILCILLTYGLYSWGQALREEGALLA	60				
Db	1	maprgcaghppspqacvcpqkmlangalagfwilcilltygliswqaleeegalla	60				
QY	61	QAGEKLEPSTTSQPHLIFILADDOQGRDVGYHGSSEIKTPTLKLAAEGVKLENYYQVP	120				
Db	61	qageklepsttsqphliladdqgrdvgyhgseiktptldklaaegvklenyyvqp	120				
QY	121	ICTPSSQFITGKYQIHTGLOHSIIRTPQNCPLDPLDNATLPQKLKEVGYSTHMVGKWHLG	180				
Db	121	ictpssqfittgkyqihgtlqhsliirtpqncpldnatlpqklkevgysthmvgkwhlg	180				
QY	181	FYRKECMPTRRGFTDFGSLGSDYYTHYKCDSPGCMGYDLYENDNAAWPDNGIYSTQ	240				
Db	181	fyrkecmptrrgfdtffgslgsgdythykcdspgcmgydyendnaawdydinglystq	240				
QY	241	MYTRVQVQIILASHNTPKPIFLYIAYQAVHSPLOAPGRYFEHYRSIINRRRYAAMLSC	300				
Db	241	mytrvqvilashnptkpiflyiayqavhsploapgrfyehyrsiinrrryaamlsc	300				
QY	301	DEAINNVTLAKTYGFYNNSIITYSSDNGGQPTAGGSNPULGSKGTWEGGIRAVGPVH	360				
Db	301	deainnvtlaktygfynnsiityssdnggqptaggsnpulrgskgtweggiravgfvh	360				
QY	361	SPLLNKGTCKEVPVHITDNYPTLISLAEGQIDEDIQLDGYDIWETISEGURSPRVILH	420				
Db	361	spllknkgtckelvhitdwyptlislgaegidediqldgydiwetisegirsprvdlh	420				
QY	421	NIDPIYTKANGSWAACYGWTNTAIQSAIRVQHKKLLTGNPGYSDWVPQPSFSLGNPRW	480				
Db	421	nidpiytkangswaacygwtntaiqsaairvqhkklltgnpgydwvpqpsfslgnprw	480				
QY	481	HNRITTSSTGKSWLFNITADPVERVDLSNRYPGIVAKLLRRLSQFNKTAVPVRYPPKDP	540				
Db	481	hnerittstgkswlfnitadpvervdlsnrypgivakllrrlsqfnktavpvyppkdp	540				
QY	541	RSNPRUNGCV	550				
Db	541	rsnprlngcv	550				

RESULT 2

RESOLUT
AAY41701
ID AAY41701 standard: Protein: 515 AA.

AX
AC

XX
DT 07-DEC-1999 (first entry)

XX DE Human PRO708 protein sequence.

XX	Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;	
KW	probe; blood coagulation disorder; cancer; cellular adhesion disorder;	
KW	secreted protein; transmembrane protein.	
XX		
OS	Homo sapiens.	
XX		
PN	W09946281-A2.	
XX		
PD	16-SEP-1999.	
XX		
PF	08-MAR-1999;	99WO-US05028.
XX		
PR	10-MAR-1998;	98US-0077450.
PR	11-MAR-1998;	98US-0077632.
PR	11-MAR-1998;	98US-0077641.
PR	11-MAR-1998;	98US-0077649.
PR	12-MAR-1998;	98US-0077791.
PR	13-MAR-1998;	98US-0078004.
PR	17-MAR-1998;	98US-0040220.
PR	20-MAR-1998;	98US-0078886.
PR	20-MAR-1998;	98US-0078910.
PR	20-MAR-1998;	98US-0078936.
PR	20-MAR-1998;	98US-0078939.
PR	25-MAR-1998;	98US-0079294.
PR	26-MAR-1998;	98US-0079656.
PR	27-MAR-1998;	98US-0079663.
PR	27-MAR-1998;	98US-0079664.
PR	27-MAR-1998;	98US-0079689.
PR	27-MAR-1998;	98US-0079728.
PR	27-MAR-1998;	98US-0079786.
PR	30-MAR-1998;	98US-0079920.
PR	30-MAR-1998;	98US-0079923.
PR	31-MAR-1998;	98US-0080105.
PR	31-MAR-1998;	98US-0080107.
PR	31-MAR-1998;	98US-0080165.
PR	31-MAR-1998;	98US-0080194.
PR	01-APR-1998;	98US-0080327.
PR	01-APR-1998;	98US-0080328.
PR	01-APR-1998;	98US-0080333.
PR	01-APR-1998;	98US-0080334.
PR	08-APR-1998;	98US-0081049.
PR	08-APR-1998;	98US-0081070.
PR	08-APR-1998;	98US-0081071.
PR	09-APR-1998;	98US-0081195.
PR	09-APR-1998;	98US-0081203.
PR	09-APR-1998;	98US-0081229.
PR	15-APR-1998;	98US-0081817.
PR	15-APR-1998;	98US-0081838.
PR	15-APR-1998;	98US-0081952.
PR	15-APR-1998;	98US-0081955.
PR	21-APR-1998;	98US-0082568.
PR	21-APR-1998;	98US-0082569.
PR	22-APR-1998;	98US-0082700.
PR	22-APR-1998;	98US-0082704.
PR	22-APR-1998;	98US-0082804.
PR	23-APR-1998;	98US-0082767.
PR	23-APR-1998;	98US-0082796.
PR	27-APR-1998;	98US-0083336.
PR	28-APR-1998;	98US-0083322.
PR	29-APR-1998;	98US-0083392.
PR	29-APR-1998;	98US-0083495.
PR	29-APR-1998;	98US-0083496.
PR	29-APR-1998;	98US-0083499.
PR	29-APR-1998;	98US-0083500.
PR	29-APR-1998;	98US-0083545.
PR	29-APR-1998;	98US-0083554.
PR	29-APR-1998;	98US-0083558.
PR	29-APR-1998;	98US-0083559.
PR	30-APR-1998;	98US-0083742.
PR	05-MAY-1998;	98US-0084366.
PR	06-MAY-1998;	98US-0084414.
PR	06-MAY-1998;	98US-0084441.

CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the pRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
XX Sequence 515 AA;

Query Match 76.6%; Score 2307.5; DB 22; Length 515;
Best Local Similarity 93.7%; Pred. No. 1.5e-210;
Matches 434; Conservative 3; Mismatches 15; Indels 11; Gaps 4;
QY 1 MAPRCGAGHPPPSPQACVCGKMLAGWILCLTYGYSWGQALEEEEGALLA 60
Db 1 maprcgaghpqacvcpqkmlangalagwilclltygyswgaleeeegalla 60
QY 61 QAGEKLEPSTTSOPHLIFILADDQFRDVGHGSEIKTPTLDKLAAGVKLENYVQP 120
Db 61 qageklepsttsqphliladdqgfrdvghgseiktptldklaegvklennyvqp 120
QY 121 ICTPSRSQFITGKYQIHTGLQHSIIRTPQNCPLDNLTPQKLKEGYSTHMGKWHLG 180
Db 121 ictpsrsqfitgkyqihgtglqhsiiirtpqncpldnltpqklkevgysthmgvkwhlg 180
QY 181 FYRKECMPTRRGFTFPGSLGSDYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQ 240
Db 181 fyrkecmptrrgfdtfpgslgsdvythykcdspgcmgydlyendnaawdydngiystq 240
QY 241 MYTORVQOILASHNPTKPIFLYTAQAVHSPLOAPGRYFEHYRSIININRRYAAMLSC 300
Db 241 mytrqvqilashnptkpiflytaqavhsploapgrgyfeyhrysiinrrryaamlsc 300
QY 301 DEAINNVTLKTYGYFNNSIIYSSDNGGQPTAGGSNWPLRGSKGYWEGGIRAVGFVH 360
Db 301 deainnvtlalktygyfnnsiiyssdnggqptaggsnwnplrgskgyweggiravgfvh 360
QY 361 SPLKNGKGVCKEVPVHTDWMYPTLISLAEGOIDEDIQDGYDIWETISEGLRSPRV 420
Db 361 spllknkgvckelvhtdwmyptlislasegoidediqlgydiwetiseglrsprvdlh 420
QY 421 NIDPIYV--KAKNGSNAAGYIWNATIQSAIRVQHWKLLTGN 460
Db 421 nidp-ycprkmpgqqamsgs---tiqss---qpsecstgn 455

RESULT 5
AAB85482
ID AAB85482 standard; Protein; 569 AA.
XX
AC AAB85482;
DT 25-SEP-2001 (first entry)
XX
DE Human 25278 sulfatase polypeptide.
XX
KW Sulfatase; 25278 sulfatase; nootropic; neuroprotective; antibacterial;
KW antiinflammatory; vasotropic; antitumor; gene therapy; human.
OS Homo sapiens.
XX
XX WO200155411-A2.
XX
PD 02-AUG-2001.
XX
PF 31-JAN-2001; 2001WO-US03266.
XX
PR 31-JAN-2000; 2000US-0495823.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
XX
DR WPI; 2001-476214/51.

DR N-PSDB; AAH46862, AAH46866.
XX
XX Novel human sulfatase polypeptides useful for treating and diagnosing
PT sulfatase-related disorders such as cerebrovascular diseases, acute
PT meningitis, multiple sclerosis, degenerative diseases and tumor -
XX
XX Claim 9; Fig 10; 180pp; English.
XX
CC The invention provides 22438, 23553, 25278 or 26212 human sulfatase
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides
CC are useful for treating disorders involving the brain such as
CC cerebrovascular diseases, infections such as acute meningitis,
CC demyelinating diseases including multiple sclerosis, degenerative
CC diseases affecting the cerebral cortex including Alzheimer's disease
CC and Pick disease, spinocerebellar degenerations including spinocerebellar
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,
CC degenerative diseases affecting motor neurons including amyotrophic
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,
CC toxic and acquired metabolic diseases, including vitamin deficiencies,
CC and neurocutaneous syndromes (phakamatoses) including neurofibromatosis.
CC The present sequence represents a human 25278 sulfatase polypeptide.
XX
XX Sequence 569 AA;
QY 27 MGALAGFWILCLLTGYLSWGQALEEEEGALLA----QAGEKLEPSTTSOPHLIFIL 82
Db 1 mhtlsgfslvslsfyglswdwa----kpsfvadpggeage--qpsaapqpghifil 53
QY 83 ADDGFRDVGHGSEIKTPTLDKLAAGVKLENYVQPICTPSRSQFITGKYQIHTGLQ 142
Db 54 tddggyhadvghgsdiedtptldriaakgklyenyvdiqptcpsqltgyrhtglq 113
QY 143 SIIRPTQNCPLDNLTPQKLKEGYSTHMGKWHLGFKYRKECMPTRRGFTFPGSLG 202
Db 114 siirpqpncpldqvtlpqklqegysthmgvkwhlgyrkeclptrrgfdtflgslg 173
QY 203 SGDYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQMYTORVQOILASHNPTKPIFLY 262
Db 174 nvdytydncdpgvcgfdihgenavagl-sgdystmlyaqrashilashspqrlfly 232
QY 263 IAYQAVHSPLOAPGRYFEHYRSIININRRYAAMLSCIDEAINNVTLKTYGYFNNSII 322
Db 233 vafqavhtplqspreylyryrtmgnvarryaamvcmdeavrnitwalkrygfynnsvl 292
QY 323 IYSSDNGGQPTAGGSNWPLRGSKGYWEGGIRAVGFVHSPLLKNGKGTVCKEPVHITD 382
Db 293 ifssdnggqptfsggsnwnplrgktywegvgvrglghvshpilkkrqtrstalmhitdwp 352
QY 383 TLISLAEGQIDEDIQDGYDIWETISEGLRSPRVDIHLNIDPIYTKAKNGSNAAGYIWN 442
Db 353 tlvglagtttsaadglgdyvwpaisegrasprteihlnidplynhaghslegfglwn 412
QY 443 TAIQSAIRVQHWKLLTGNPGYSWVPPQSFNSLGNPRWHNERITSSTCKSVLENITADP 502
Db 413 tavqaairvgewklltgdpgygdwlpptqtlatf-pgswnlnlernasvrqavwlfnisadp 471
QY 503 YERVDLSNRPYGVIVKLLRLRSQFNKTAVPVRYPPKDPSPRNLNGG 549
Db 472 yeredlagqrpdvvtllarlaeynrtaiptyrpaenprahpfdngg 518
RESULT 6
AAU14421
ID AAU14421 standard; Protein; 289 AA.
XX
AC AAU14421;
XX
DT 24-OCT-2001 (first entry)

XX DE Human novel protein #292.
XX KW Human: novel protein; Antianaemic; osteopathic; antiinflammatory;
XX KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
XX KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
XX KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
XX KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
XX KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX KW tissue regeneration; immune disorder.
XX OS Homo sapiens.
XX PN WO200155437-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US02623.
XX PR 25-JAN-2000; 2000US-0491404.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-451939/48.
XX DR N-PSDB; AAS22726.
XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
XX PT nervous system disorders, and for regenerating bone and cartilage -
XX PS Example 4; Page 809-810; 894pp; English.

CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence represents a protein of the invention.

XX SQ Sequence 289 AA;
Query Match 51.9%; Score 1562; DB 22; Length 289;
Best Local Similarity 100.0%; Pred. No. 5.8e-140;
Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 MIANGALAGFWILCLLYGYLSWQALKEEEEGALLAQAGEKLEPSTSTSQPHLIFILA 83
DB 1 mlangalagfwilcllygylswgaleeeegallagakelepststsqphliffila 60
QY 84 DDQCFRDVGYHSGSEIKPTLDKLAAGVKLENYYVQICTPSRSQFTGKYIHTGLQHS 143
DB 61 ddqgfrdvgyhsgseikptldklaagvklennyvqictpsrsgftgkyihtgqlghs 120

QY 144 IIRPTQPNCLPLDNATLPQKLKEVGYSTHMYGKWHLGEYRKECMPTRRGFDTFGSLGGS 203
DB 121 iirptqpnclpldnatlpqklkevgysthmvgkwhlgfyркеcmptrrgfdtffgsligs 180
QY 204 GDYTHYKCDSPGCGYDLYENDNAANDYNGIYSTQMYTORVQOILASHNPTKPIFLYI 263
DB 181 gdytyhykcdspgcmgydlyendnaawdynglystqmytrvqoilashnptkpiilyi 240
QY 264 AYQAVHSPLOQPGYFHEHYRSIININRRRYAAMLSCDLDEAINNVTLALK 312
DB 241 ayqavhsploqpgryfeyhryslinlnrrryaamisclddeainnvtlalk 289

RESULT 7
AAB51184
ID AAB51184 standard; Protein; 533 AA.
XX AC AAB51184;
XX DT 21-MAR-2001 (first entry)
XX DE Human sulfatase protein B SEQ ID NO:13.
XX KW Human; iduronate 2-sulfatase; IDS; glycosylated; gene therapy;
KW mucopolysaccharidosis inhibitor; IDS deficiency disorder;
KW Hunter syndrome; mucopolysaccharidosis type II.
XX OS Homo sapiens.
XX PN US6153188-A.
XX PD 28-NOV-2000.
XX PF 12-FEB-1999; 99US-0249003.
XX PR 17-DEC-1992; 92US-0991973.
XX PR 28-NOV-1994; 94US-0345212.
XX PR 12-NOV-1991; 91US-0790362.
XX PA (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.
XX PI Wilson PJ, Anson DS, Occhiodoro T, Bielicki J, Clements PR;
PI Hopwood JJ, Morris CP;
XX DR WPI; 2001-060075/07.

XX New highly glycosylated recombinant human iduronate 2-sulfatase (IDS)
XX useful for diagnosing or treating subjects suspected of having or
XX suffering from IDS deficiency disorders, e.g. Hunter syndrome
XX (mucopolysaccharidosis-II) -
XX Example 1; Fig 3; 53pp; English.
XX The present invention describes a recombinant human iduronate 2-sulfatase
XX (IDS). The recombinant IDS is more highly glycosylated than the naturally
XX occurring enzyme isolated from human tissue. The recombinant human IDS
XX can be produced in Chinese Hamster Ovary (CHO) cells or in a human cell.
XX The recombinant IDS comprises a fusion protein. It is a
XX mucopolysaccharidosis inhibitor and can be used in gene therapy.
XX The recombinant IDS is useful in treating and diagnosing subjects
XX suffering from or suspected of having IDS deficiency disorders, e.g.
XX Hunter syndrome (mucopolysaccharidosis type II). The present sequence
XX represents a human sulfatase which is used in comparison with the
XX IDS sequence in an example from the present invention.

XX SQ Sequence 533 AA;

Query Match 47.0%; Score 1415.5; DB 22; Length 533;
Best Local Similarity 49.1%; Pred. No. 1.4e-125;
Matches 278; Conservative 80; Mismatches 131; Indels 57; Gaps 8;

dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; diabetes; severe combined immunodeficiency; allergic rhinitis; osteoporosis; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder.

OS Homo sapiens.
XX WO200153455-A2.
XX 26-JUL-2001.
XX
XX 22-DEC-2000; 2000WO-US35017.
XX
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-457603/49.
DR N-PSDB; AAH99356.
XX

Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
Claim 20; Page 200; 1217pp; English.

XX
XX AAH99166 to AAH99904 encode the human proteins given in AAH25225 to
CC AAH25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.

XX Sequence 187 AA;

Query Match 32.8%; Score 989; DB 22; Length 187;
Best Local Similarity 99.4%; Pred. No. 1e-85;
Matches 179; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 133 KYQHTGLQHSIIIRPTQNCPLDNLATLPQKLKEVGYSTHMVGKHLGFYKCEMPTRRG 192
Db :|||||
8 ryqhtqlsniirptqncpldnlatlpqklkevgysthmvgkwhlgyfyrkcmptrrg 67

QY 193 FDTFFGSLGSGDYTHYKDSPGMCGYDLYENDNAADYDNGIYSTOMYTORVQOITLAS 252
Db :|||||
68 fdtffgslsgsdythykcdspgmcydlyendnaadynyngiysctmyqtrvqqilas 127

QY 253 HNPTKPIFLYTAQVHSLQAPGRYFEHYRSIINRRRYAAMLSCIDELAINNVTLALK 312

Db 128 hnptkpiflytaqvahslqapgrgyfehyrsiinrrryaamlscideainnvtlalk 187
:|||||

RESULT 10
ABBG3962
ID ABBG3962 standard; Protein; 542 AA.

XX ABBG3962;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 18678.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

XX N-PSDB; ABL08065.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

PS Disclosure; SEQ ID NO 18678; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLL16176-ABLL30511), expressed DNA
CC sequences (ABLL01840-ABLL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 542 AA;

Query Match 30.5%; Score 920; DB 22; Length 542;
Best Local Similarity 38.1%; Pred. No. 2.2e-78;
Matches 213; Conservative 79; Mismatches 169; Indels 98; Gaps 15;

QY 34 WIGLLLYGYLSWGQALEEBEEGALLAQAGKLEPSTSTSQPHLIFILADDDQGRDVG 93
Db :|||||
5 wlicell----lpiidaeevek-----spakpnifiladldgndvfg 43

QY 94 HGS-EIKTPTLDKLAAGVKLENYYVQICTPSRSQFTGKYQIHTGLQHSIIIRPTQNC 152
Db :|||||
44 hgsaieptpidaLaysgiiInryyvapictpsrsaitmkyphthgmqhtvlyaaepg 103

QY 153 LPLDNATLPQKLKEVGYSTHMVGKHLGFYKCEMPTRRGDTFFGSLGSGDYTHYK 212
Db :|||||
104 lpleekilpdylnelgysthshagkwhlghklytlyrgrfsshwgldmngt----- 156

QY 213 DSPGMCGYDLYENDNAADYDNGIYSTOMYTORVQOITLASHNPTK-PIFLYTAQVHSL 270

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0246167.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-465566/50.
DR
XX N-PSDB; AAS41205.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases
XX
XX Claim 11; SEQ ID No 1331; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 128 AA;

Query Match 20.0%; Score 603.5; DB 22; Length 128;
Best Local Similarity 94.2%; Pred. No. 2.9e-49;
Matches 113; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 90 DVGYHGSEIKTPTLDKLAAGVKLENYYVQICTPSRSQFTGKYQIHTGLQHSIRPTQ 149
|||||
Db 6 dvyghgseiktptldklaaegvklennyvvpictpsrsqftgkyqhtglqhsirptq 65
|||||

Qy 150 PNCPLDNLATLPQKLKEVGYSTHMGVKGWHLGFGYRKCEMPTRRGDFTFGSLGSGDYVT 208
|||||
Db 66 pncpldnlatlpgklkevgysthmgvkgwhlgfyrkcepmtrgrfdttffgfgkgwllt 125
|||||

RESULT 14
ABB63259
ID ABB63259 standard; Protein: 486 AA.
XX
XX ABB63259;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX

Db 72 plsspsraglltgmprftgir-swipsgkdvalgrneltianlllkaggydtammgklhl 130
QY 180 ---GFYRKECMPTRRGEDTFFGSLGSGDYTHYKCDSPGCMGYDLYENDNAAWDYDNGI 236
Db 131 naggdrtddqqaqmgfdy---slantagfvtdatld-----nakerprygmvyptgw 180
QY 237 YSTOMYTORVQOI-----LASHNPTKPIFLYIAYQAVHSPLQAPGRYFEHYRS 284
Db 181 lrngqptpradkmsgeyvssevvnwldnkkdkpfflyvaftevhspkplaskyldmysq 240
QY 285 IININRRR-----YAAWLSCLDEAINNVTLALATYGFYNNNSIIYS 325
Db 241 ymsayqkqhpdlfygdwadkpwrgvgeyyanisyldgaqvkvlckikamgeedhtlvi 300
QY 326 SDNGG-----QPTAGGSNNWPLRGSKGTWEGGIRAVGFV-----HSPLLKNKGTVC 372
Db 301 sdngpvtrearkvyelnlagetdglgrkdnlweggirvpailkygkhlp-----qgmvsd 356
QY 373 EPVHITDWTPLISLAEGQIDEDQLDGYDIWETISEGLRSPRVDILHNIDPIYTKAKNG 432
Db 357 tpvygldwptlakmmnfkiptdrtdfgeslvpvleqkalkrekplifgidmpfqddptd 416
QY 433 SWAAGYGIWNTAIQSAIRVOHWKLLTGNPGYSDWVPPQSFSNLGPNRWHNERITSS 492
Db 417 ew-----alrdgdwkm-----iidrnknk 435
QY 493 VWLFNITADPYERVDLSNRYPGIVKLLRLRSQFNKTAV 531
Db 436 kylvnlksdryetlnlllgkxpdiyekmqygfiky-kt 473

Search completed: July 23, 2002, 19:02:50
Job time: 6936 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:04:14 ; Search time 22.71 seconds
(without alignments)
2327.131 Million cell updates/sec

Title: US-09-495-823-7

Perfect score: 550

Sequence: 1 MAPRGACGHPPPSPQACVC.....VPVRYPPKDRSPRLNGGV 550

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	2.0	473	I54210	N-acetylglactosam
2	10	1.8	289	AI0374	probable thioredox
3	10	1.8	533	KJHUB	N-acetylglactosam
4	10	1.8	535	A44475	N-acetylglactosam
5	8	1.5	176	S76001	phosphoribosylamin
6	8	1.5	219	A23385	two-component resp
7	8	1.5	330	A83230	3-oxoacyl-[acyl-ca
8	8	1.5	381	AD1113	hypothetical prote
9	8	1.5	381	AH1474	hypothetical prote
10	8	1.5	452	T29618	hypothetical prote
11	8	1.5	551	S07089	arylsulfatase (EC
12	8	1.5	551	S01793	arylsulfatase (EC
13	8	1.5	567	A37362	arylsulfatase (EC
14	8	1.5	647	S61973	hypothetical prote
15	8	1.5	1216	AH1335	pyruvate-flavodoxi
16	8	1.5	1420	B57062	SRB9 protein - yea
17	7	1.3	82	AD2287	hypothetical prote
18	7	1.3	108	AG0271	translation initia
19	7	1.3	114	E70327	hypothetical prote
20	7	1.3	123	H72468	hypothetical prote
21	7	1.3	123	H84014	hypothetical prote
22	7	1.3	155	AI2450	hypothetical prote
23	7	1.3	159	AH1524	hypothetical prote
24	7	1.3	180	C39415	fimbrial protein s
25	7	1.3	198	T45541	septum-promoting G
26	7	1.3	200	T49590	hypothetical prote
27	7	1.3	220	C83292	probable glutathio
28	7	1.3	229	F84685	hypothetical prote
29	7	1.3	232	AH0252	conserved hypothet

30	7	1.3	237	2	AC07558	precorrin-2 C20-me
31	7	1.3	238	1	Q0BE74	US4 protein - huma
32	7	1.3	241	2	T20432	hypothetical prote
33	7	1.3	248	2	AI2175	hypothetical prote
34	7	1.3	277	2	AD3348	creatininase (EC 3
35	7	1.3	278	2	G81148	pantoate-beta-ala
36	7	1.3	286	2	E91262	probable aldolase
37	7	1.3	286	2	H86102	hypothetical prote
38	7	1.3	286	2	AI2390	hypothetical prote
39	7	1.3	294	2	AB0115	probable carbon-ni
40	7	1.3	298	2	JC7568	kidney inhibitor o
41	7	1.3	301	2	AF2004	33kD chaperonin, h
42	7	1.3	302	2	S75112	hypothetical prote
43	7	1.3	303	2	T15835	hypothetical prote
44	7	1.3	304	2	B83951	tRNA pseudouridine
45	7	1.3	319	2	A98140	glycosyl transfera

ALIGNMENTS

RESULT 1

I54210

N-acetylglactosamine-4-sulfatase (EC 3.1.6.12) precursor - rat (fragment)
N;Alternate names: arylsulfatase (EC 3.1.6.1) B [misidentification]
C;Species: Rattus norvegicus (Norway rat)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 02-Jun-2000
C;Accession: I54210

R;Kunieda, T.

Genomics 29, 582-587, 1995

A;Title: Mucopolysaccharidosis type VI in rats: Isolation of cDNAs encoding arylsulfatase
A;Reference number: I54210; MUID:96121368

A;Accession: I54210

A;Status: preliminary; translated from GB/EMBL/DBDJB

A;Molecule type: mRNA

A;Residues: 1-473 <RES>

A;Cross-references: GB:D49434; NID:g1065603; PIDN:BAA08412.1; PID:g1089794

C;Genetics:

A;Gene: ARSB

C;Superfamily: animal sulfatase

C;Keywords: sulfuric ester hydrolase

F;31/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 2.0% Score 11; DB 2; Length 473;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 STHMVGRKWHLG 180

|||||

Db 79 STHMVGRKWHLG 89

RESULT 2

AI0374

probable thioredoxin YP03082 [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C;Accession: AI0374

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AI0374

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-289 <KUR>

A;Cross-references: GB:AL590842; PIDN:CAC92324.1; PID:g15981035; GSPDB:GN00175

C;Genetics:

A;Gene: YP03082

A;Cross-references: GB:S48472; NID:g258855; PIDN:AAB23941.1; PID:g258856
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:117976)
C;Comment: This enzyme is frequently misidentified as EC 3.1.6.1.
C;Function:
A;Description: hydrolyzes N-acetylgalactosamine-4-sulfate units in chondroitin sulfate A
C;Superfamily: animal sulfatase
C;Keywords: glycoprotein; lysosomal storage disease; lysosome; sulfuric ester hydrolase
F;1-42/Domain: signal sequence #status predicted <SIG>
F;43-425/Product: alpha chain #status predicted <MAT>
F;426-467/Product: gamma chain #status predicted <MAT>
F;468-535/Product: beta chain #status predicted <MAT>
F;93/Modified site: 3-oxoalanine (Cys) #status predicted
F;190,281,428,460/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.8%; Score 10; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 THMVGKWHLG 180
|||||
Db 142 THMVGKWHLG 151

RESULT 5
S76001
phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21) catalytic chain - Synecocystis s
N;Alternate names: protein sll0901
C;Species: Synecocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S76001
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76001
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-176 <KAN>
A;Cross-references: EMBL:D64006; GB:AB001339; NID:g1001291; PIDN:BAA10848.1; PID:g100136
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Gene: pure
C;Superfamily: phosphoribosylaminoimidazole carboxylase catalytic chain; phosphoribosyla
C;Keywords: carbon-carbon lyase; carboxy-lyase; purine nucleotide biosynthesis
F;6-139/Domain: phosphoribosylaminoimidazole carboxylase catalytic chain homology <PCC>

Query Match 1.5%; Score 8; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 QILASHNP 255
|||||
Db 131 QILASHNP 138

RESULT 6
AC2385
two-component response regulator all4635 [imported] - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C;Accession: AC2385
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AC2385
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA076334.1; PID:g17133772; GSPDB:GN001179
A;Experimental source: strain PCC 7120
C;Genetics:
C;Superfamily: regulatory protein comA; response regulator homology

Query Match 1.5%; Score 8; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 PGIVKKLL 520
|||||
Db 127 PGIVKKLL 134

RESULT 7
AB3230
3-oxoacyl-[acyl-carrier-protein] synthase (EC 2.3.1.41) III PA3333 [similarity] - Pse
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 02-Feb-2001
C;Accession: AB3230
R;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Miziochuchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: AB2950; MUID:20437337
A;Accession: AB3230
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-330 <STO>
A;Cross-references: GB:AE004755; GB:AE004091; NID:g9949456; PIDN:AAG06721.1; GSPDB:GN
A;Experimental source: strain PA01
C;Genetics:
A;Gene: fabH2; PA3333
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase III
C;Keywords: acyltransferase

Query Match 1.5%; Score 8; DB 2; Length 330;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 EEEGALLA 60
|||||
Db 171 EEEGALLA 178

RESULT 8
AD1113
hypothetical protein lmo0307 [imported] - Listeria monocytogenes (strain EGD-e)
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AD1113
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AD1113
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00834.1; PID:g16409672; GSPDB:GN00177
A;Experimental source: strain EGD-e

C;Genetics:
A;Gene: lmo0307

Query Match 1.5%; Score 8; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 315 GFYNNII 322
|||||

Db 181 GFYNNII 188

RESULT 9

hypothetical protein lin0335 [imported] - *Listeria innocua* (strain Clp11262)
C;Species: *Listeria innocua*
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AH1474
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1474
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-381 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95568.1; PID:g16412764; GSPDB:GN00178
A;Experimental source: strain Clp11262
C;Genetics:
A;Gene: lin0335

Query Match 1.5%; Score 8; DB 2; Length 381;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 GFYNNII 322
|||||

Db 181 GFYNNII 188

RESULT 10

hypothetical protein D1014.1 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29618
R;Du, Z.; Leinbac, D.
submitted to the EMBL Data Library, March 1996
A;Description: The sequence of *C. elegans* cosmid D1014.
A;Reference number: Z20652
A;Accession: T29618
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-452 <DUZ>
A;Cross-references: EMBL:U53180; PIDN:AAA96290.1; GSPDB:GN00023; CESP:D1014.1
A;Experimental source: strain Bristol N2; clone D1014
C;Genetics:
A;Gene: CESP:D1014.1
A;Map position: 5
A;Introns: 256/3; 336/2; 376/3; 405/3

Query Match 1.5%; Score 8; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180

Db 134 MVGKWHLG 141
|||||

RESULT 11

arylsulfatase (EC 3.1.6.1) - sea urchin (*Lytechinus pictus*)
C;Species: *Lytechinus pictus* (painted urchin)
C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 26-Aug-1999
C;Accession: S07089
R;Yamada, K.; Akasaka, K.; Shimada, H.
Eur. J. Biochem. 186, 405-410, 1989
A;Title: Structure of sea-urchin arylsulfatase gene.
A;Reference number: S07089; MUID:90092130

A;Accession: S07089
A;Molecule type: DNA
A;Residues: 1-551 <YAM>
A;Cross-references: EMBL:X16679
C;Genetics:

A;Gene: Ars
A;Introns: 103/2; 160/1; 197/3; 272/1; 327/1
C;Superfamily: animal sulfatase
C;Keywords: sulfuric ester hydrolase
F:100/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 1.5%; Score 8; DB 2; Length 551;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180
|||||

Db 153 MVGKWHLG 160

RESULT 12

arylsulfatase (EC 3.1.6.1) precursor - sea urchin (*Hemicentrotus pulcherrimus*)
C;Species: *Hemicentrotus pulcherrimus*
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1991 #text_change 26-Aug-1999
C;Accession: S01793
R;Sasaki, H.; Yamada, K.; Akasaka, K.; Kawasaki, H.; Suzuki, K.; Saito, A.; Sato, M.;
Eur. J. Biochem. 177, 9-13, 1988
A;Title: cDNA cloning, nucleotide sequence and expression of the gene for arylsulfatase
A;Reference number: S01793; MUID:89030699
A;Accession: S01793
A;Molecule type: mRNA
A;Residues: 1-551 <SAS>
A;Cross-references: EMBL:X17015; MID:g9432; PID:g9433
A;Note: part of this sequence was confirmed by protein sequencing
A;Note: the authors translated the codon CAC for residue 61 as Asp
C;Superfamily: animal sulfatase
C;Keywords: sulfuric ester hydrolase
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-551/Product: arylsulfatase #status predicted <MAT>
F:100/Modified site: 3-oxoalanine (Cys) #status predicted

A;Accession: S01793
A;Molecule type: mRNA
A;Residues: 1-551 <SAS>
A;Cross-references: EMBL:X17015; MID:g9432; PID:g9433
A;Note: part of this sequence was confirmed by protein sequencing
A;Note: the authors translated the codon CAC for residue 61 as Asp
C;Superfamily: animal sulfatase
C;Keywords: sulfuric ester hydrolase
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-551/Product: arylsulfatase #status predicted <MAT>
F:100/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 1.5%; Score 8; DB 2; Length 551;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180
|||||

Db 153 MVGKWHLG 160

RESULT 13

arylsulfatase (EC 3.1.6.1) precursor - sea urchin (*Strongylocentrotus purpuratus*)
C;Species: *Strongylocentrotus purpuratus* (purple urchin)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Nov-1999
C;Accession: A37362

R:Yang, Q.; Angerer, L.M.; Angerer, R.C.

Dev. Biol. 135, 53-65, 1989

A:Title: Structure and tissue-specific developmental expression of a sea urchin arylsulphatase

A:Reference number: A37362; MUID:89357267

A:Accession: A37362

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-567 <YAN>

A:Cross-references: GB:M28404; GB:M25815; NID:g161440; PID:g161441

C:Superfamily: animal sulfatase

C:Keywords: sulfuric ester hydrolase

F:115/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 1.5%; Score 8; DB 2; Length 567;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180

|||||||

Db 168 MVGKWHLG 175

RESULT 14

S61973

hypothetical protein YPL093w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein LPG15w

C:Species: Saccharomyces cerevisiae

C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 05-Nov-1999

C:Accession: S61973

R:Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.; Voigt, M.; et al. Submitted to the EMBL Data Library, December 1995

A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A:Reference number: S61959

A:Accession: S61973

A:Molecule type: DNA

A:Residues: 1-647 <WAN>

A:Cross-references: EMBL:U43281; NID:g1151218; PID:g1151233; GSPDB:GN00016; MIPS:YPL093w

C:Genetics:

A:Gene: MIPS:YPL093w

A:Map position: 16L

Query Match

1.5%; Score 8; DB 2; Length 647;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 QALEEEEE 55

|||||||

Db 454 QALEEEEE 461

RESULT 15

AH1535

pyruvate-flavodoxin oxidoreductase homolog nifJ [imported] - Listeria innocua (strain C1)

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AH1535

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, H.; et al. Submitted to the EMBL Data Library, December 1995

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AH1535

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1216 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC96056.1; PID:g16413275; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: nifJ

C:Superfamily: pyruvate (flavodoxin) dehydrogenase; ferredoxin 2[4Fe-4S] homology

Query Match 1.5%; Score 8; DB 2; Length 1216;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 ETISEGLR 412

|||||||

Db 902 ETISEGLR 909

Search completed: July 23, 2002, 19:09:17

Job time: 303 sec

Result No.	Query Match	Score	8			ID	Description
			Length	DB			
1	11	2.0	473	1	ARSB_RAT	P50430 rattus norv	
2	10	1.8	285	1	ARSB_MOUSE	P50429 mus musculu	
3	10	1.8	533	1	ARSB_HUMAN	P15848 homo sapien	
4	10	1.8	535	1	ARSB_FELCA	P33727 felis silve	
5	8	1.5	176	1	PUR6_SYNT3	Q55498 synechocyst	
6	8	1.5	551	1	ARS_HEMPU	P14000 hemiceotrop	
7	8	1.5	567	1	ARS_STRPU	P50473 strongyloce	
8	8	1.5	647	1	NOG1_YEAST	Q02892 saccharomyc	
9	8	1.5	1420	1	SRB9_YEAST	P38931 saccharomyc	
10	7	1.3	237	1	CBIL_SALTY	Q05593 salmonella	
11	7	1.3	238	1	VGLG_HSV11	P06484 herpes simp	
12	7	1.3	252	1	PSA3_ACACA	P90513 acanthamoeb	
13	7	1.3	278	1	PANC_NEIMB	P75036 neisseria m	
14	7	1.3	302	1	HSLO_SYNT3	P73910 synechocyst	
15	7	1.3	304	1	TRUB_BACHD	Q9ka80 bacillus ha	
16	7	1.3	339	1	XAP5_HUMAN	Q14320 homo sapien	
17	7	1.3	387	1	RECF_CAUCR	P49998 caulobacter	
18	7	1.3	391	1	MUCB_HUMAN	P04220 homo sapien	
19	7	1.3	397	1	MM16_MYCTU	Q10773 mycobacteri	
20	7	1.3	449	1	COMB_STRPN	P36498 streptococc	
21	7	1.3	454	1	MUC_HUMAN	P01871 homo sapien	
22	7	1.3	458	1	C6ST_CHICK	Q92179 gallus gall	
23	7	1.3	464	1	ARS_KLEAE	P20713 klebsiella	
24	7	1.3	470	1	EGR2_MOUSE	Q08152 mus musculu	
25	7	1.3	472	1	C901_ARATH	Q42569 arabidopsis	
26	7	1.3	489	1	YHIP_ECOLI	P36837 escherichia	
27	7	1.3	491	1	VIE1_HCMVA	P13202 human cytom	
28	7	1.3	491	1	VIE1_HCMVT	P03169 human cytom	
29	7	1.3	505	1	A1NX_RAT	P23565 rattus norv	
30	7	1.3	508	1	CP77_CHICK	P12394 gallus gall	
31	7	1.3	522	1	GAGS_HUMAN	P34059 homo sapien	
32	7	1.3	537	1	YRR2_EBV	P03210 epstein-bar	
33	7	1.3	550	1	LIGA_MOUSE	Q61211 mus musculu	

FT CARBOHYD 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 473 AA; 53320 MW; 4E114C923A24AF9F CRC64;

Query Match 2.0%; Score 11; DB 1; Length 473;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 STMVGVKWHLG 180
|||||
DB 79 STMVGVKWHLG 89

RESULT 2
ARSB_MOUSE
ID ARSB_MOUSE STANDARD; PRT; 285 AA.
AC P50429;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Arylsulfatase B (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-4-sulfatase) (G4S) (Fragments).
DE 4-sulfatase) (G4S) (Fragments).
GN ARSB OR AS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=92241876; PubMed=1572648;
RA Grompe M., Pieretti M., Caskey C.T., Ballabio A.;
RT "The sulfatase gene family: cross-species PCR cloning using the MOPAC technique.";
RL Genomics 12:755-760(1992).
RN [2]
RP SEQUENCE OF 34-285 FROM N.A.
RX MEDLINE=96323200; PubMed=8710849;
RA Evers M., Saftig P., Schmidt P., Hafner A., McLothlin D.B.,
RA Schmahl W., Hess B., von Figura K., Peters C.W.B.;
RT "Targeted disruption of the arylsulfatase B gene results in mice resembling the phenotype of mucopolysaccharidosis VI.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:8214-8219(1996).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and dermatan sulfate.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M82877; AAA37261.1; .
CC EMBL; X92096; CAA63067.1; .
CC HSP; P15848; IFSU
CC MGD; MGI:88075; Asi-s.
CC InterPro: IPR000917; Sulfatase.
CC Pfam; PF00884; Sulfatase; 1.
CC PROSITE; PS00523; SULFATASE_1; PARTIAL.
CC PROSITE; PS00149; SULFATASE_2; 1.
KW Hydrolase; Glycoprotein; Lysosome.
FT NON_TER 1 1
FT NON_CONS 33 34
FT ACT_SITE 45 45 POTENTIAL.

FT CARBOHYD 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 285 285
SQ SEQUENCE 285 AA; 31727 MW; 36BA5B98EA0770C9 CRC64;

Query Match 1.8%; Score 10; DB 1; Length 285;
Best Local Similarity 100.0%; Pred. No. 0.073;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 THMVGKWHLG 180
|||||
DB 38 THMVGKWHLG 47

RESULT 3
ARSB_HUMAN
ID ARSB_HUMAN STANDARD; PRT; 533 AA.
AC P15848;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Arylsulfatase B precursor (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-4-sulfatase) (G4S).
GN ARSB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90153994; PubMed=2303452;
RA Peters C., Schmidt B., Rommerskirch W., Rupp K., Zuehlendorf M.,
RA Vangron M., Meyer H.E., Pohlmann R., von Figura K.;
RT "Phylogenetic conservation of arylsulfatases. cDNA cloning and expression of human arylsulfatase B.";
RL J. Biol. Chem. 265:3374-3381(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90152677; PubMed=1968043;
RA Schuchman E.H., Jackson C.E., Desnick R.J.;
RT "Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-length cDNA, and regions of amino acid identity with arylsulfatases A and C.";
RL Genomics 6:149-158(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93332648; PubMed=7687847;
RA Modaresi S., Rupp K., von Figura K., Peters C.;
RT "Structure of the human arylsulfatase B gene.";
RL Biol. Chem. Hoppe-Seyler 374:327-335(1993).
RN [4]
RP SEQUENCE OF 1-104 FROM N.A.
RX MEDLINE=92028992; PubMed=1930244;
RA Litjens T., Morris C.P., Gibson G.J., Beckmann K.R., Hopwood J.J.;
RT "Human N-acetylgalactosamine-4-sulphatase: protein maturation and isolation of genomic clones.";
RL Biochem. Int. 24:209-215(1991).
RN [5]
RP 2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.
RX MEDLINE=95354208; PubMed=7628016;
RA Schmidt B., Selmer T., Ingendoh A., von Figura K.;
RT "A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency.";
RL Cell 82:271-278(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=97184692; PubMed=9032078;
RA Bond C.S., Clements P.R., Ashby S.J., Collyer C.A., Harrop S.J., Hopwood J.J., Guss J.M.;
RT "Structure of a human lysosomal sulfatase.";
RN

RL Structure 5:277-289(1997).
RN [7]
RP VARIANTS MPS-VI ARG-117; PRO-236 AND TYR-405.
RX MEDLINE-92197625; PubMed-1550123;
RA Jin W.-D., Jackson C.E., Desnick R.J., Schuchman E.H.;
RT "Mucopolysaccharidosis type VI: identification of three mutations in
the arylsulfatase B gene of patients with the severe and mild
phenotypes provides molecular evidence for genetic heterogeneity.";
RL Am. J. Hum. Genet. 50:795-800(1992).
RN [8]
RP VARIANT MPS-VI VAL-137, AND VARIANT MET-376.
RX MEDLINE-92042029; PubMed-1718978;
RA Wicker G., Prill V., Brooks D., Gibson G., Hopwood J.,
von Figura K., Peters C.;
RT "Mucopolysaccharidosis VI (Maroteaux-Lamy syndrome). An intermediate
clinical phenotype caused by substitution of valine for glycine at
position 137 of arylsulfatase B.";
RL J. Biol. Chem. 266:21386-21391(1991).
RN [9]
RP VARIANTS MPS-VI MET-92; GLN-95; CYS-210; PRO-393 AND PRO-498.
RX MEDLINE-96213747; PubMed-8651289;
RA Litjens T., Brooks D.A., Peters C., Gibson G.J., Hopwood J.J.;
RT "Identification, expression, and biochemical characterization of N-
acetylglucosamine-4-sulfatase mutations and relationship with
clinical phenotype in MPS-VI patients.";
RL Am. J. Hum. Genet. 58:1127-1134(1996).
RN [10]
RP VARIANTS MPS-VI TRP-152 AND GLN-160.
RX MEDLINE-94171224; PubMed-8125475;
RA Voskoboeva E., Isbrandt D., von Figura K., Krasnopolskaya X.,
Peters C.;
RT "Four novel mutant alleles of the arylsulfatase B gene in two
patients with intermediate form of mucopolysaccharidosis VI
(Maroteaux-Lamy syndrome).";
RL Hum. Genet. 93:259-264(1994).
RN [11]
RP VARIANT MPS-IV ARG-302.
RA Villani G.R.D., Balzano N., di Natale P.;
RT "Two novel mutations of the arylsulfatase B gene in two Italian
patients with severe form of mucopolysaccharidosis.";
RL Hum. Mutat. 11:410-410(1998).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-
acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and
dermatan sulfate.
CC -I- SUBUNIT: MONOMER.
CC -I- SUBCELLULAR LOCATION: Lysosomal.
CC -I- DISEASE: DEFECTS IN ARSB ARE THE CAUSE OF MAROTEAUX-LAMY
SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).
CC THIS DISEASE IS CHARACTERIZED BY THE ACCUMULATION OF DERMATAN
SULFATE IN LYOSOMES. CLINICAL FEATURES CAN INCLUDE ABNORMAL
GROWTH, SHORT STATURE, STIFF JOINTS, SKELETAL MALFORMATIONS,
CORNEAL CLOUDING, HEPATOSPLENOMEGALY, AND CARDIAC ABNORMALITIES.
CC A WIDE VARIATION IN CLINICAL SEVERITY IS OBSERVED.
CC -I- DISEASE: MULTIPLE SULFATASE DEFICIENCY (MSD) IS A DISORDER THAT
COMBINES FEATURES OF METACHROMATIC LEUKODYSTROPHY AND OF
MUCOPOLYSACCHARIDOSIS; IT IS CHARACTERIZED BY A DECREASED
ACTIVITY OF ALL KNOWN SULFATASES. IT SEEMS TO BE CAUSED FROM THE
LACK OF POST-TRANSLATIONAL MODIFICATION OF A CYSTEINE INTO 2-
AMINO-3-OXOPROPIONIC ACID.
CC -I- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; J05225; AAA51784.1; -;
DR EMBL; M32373; AAA51779.1; -;
DR EMBL; X72735; CAA51272.1; -;
DR EMBL; X72736; CAA51272.1; JOINED.

DR EMBL; X72737; CAA51272.1; JOINED.
DR EMBL; X72738; CAA51272.1; JOINED.
DR EMBL; X72739; CAA51272.1; JOINED.
DR EMBL; X72740; CAA51272.1; JOINED.
DR EMBL; X72741; CAA51272.1; JOINED.
DR EMBL; X72742; CAA51272.1; JOINED.
DR EMBL; S57777; AAB19988.1; -;
DR PIR; A35078; A35078.
DR PDB; 1FSU; 04-FEB-98.
DR MIM; 253200; -;
DR MIM; 272200; -;
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Hydrolase; Signal; Glycoprotein; Lysosome; Mucopolysaccharidosis;
Disease mutation; Polymorphism; 3D-structure;
FT SIGNAL 1 36
OR 38 (POTENTIAL).
FT CHAIN 37 533
ARYLSULFATASE B.
FT ACT_SITE 147 147
POTENTIAL.
FT MOD_RES 91 91
2-AMINO-3-OXOPROPIONIC ACID.
FT DISULFID 117 521
FT DISULFID 121 155
FT DISULFID 181 192
FT DISULFID 405 447
FT CARBOHYD 188 188
FT CARBOHYD 279 279
FT CARBOHYD 291 291
FT CARBOHYD 366 366
FT CARBOHYD 426 426
FT CARBOHYD 458 458
FT VARIANT 92 92
N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 95 95
R -> Q (IN MPS-VI; MILD/SEVERE FORM).
FT VARIANT 117 117
C -> R (IN MPS-VI; SEVERE FORM).
FT VARIANT 137 137
G -> V (IN MPS-VI; INTERMEDIATE FORM).
FT VARIANT 152 152
R -> W (IN MPS-VI; INTERMEDIATE FORM).
FT VARIANT 160 160
R -> Q (IN MPS-VI; INTERMEDIATE FORM).
FT VARIANT 210 210
Y -> C (IN MPS-VI; MILD/INTERMEDIATE).
FT VARIANT 236 236
L -> P (IN MPS-VI; MILD FORM).
FT VARIANT 302 302
G -> R (IN MPS-VI; SEVERE FORM).
FT VARIANT 376 376
V -> M.
FT VARIANT 393 393
H -> P (IN MPS-VI; MILD/SEVERE FORM).
FT VARIANT 405 405
C -> Y (IN MPS-VI; MILD FORM).
FT VARIANT 498 498
L -> P (IN MPS-VI; MILD/SEVERE FORM).
FT CONFLICT 358 358
V -> M (IN REF. 3).
FT SEQUENCE 533 AA; 59687 MW; 5983FB691C4789A CRC64;

Query Match 1.8%; Score 10; DB 1; Length 533;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 THMVGKWHLG 180
|||||||
Db 140 THMVGKWHLG 149

RESULT 4
ARSB_FELCA STANDARD; PRY; 535 AA.
ID ARSB_FELCA

```
AC P33727;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Arylsulfatase B precursor (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-
DE 4-sulfatase) (G4S).
GN ARSB.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93052342; PubMed=1427856;
RA Jackson C.E., Yuhki N., Desnick R.J., Haskins M.E., O'Brien S.J.,
RA Schuchman E.H.;
RT "Feline arylsulfatase B (ARSB): isolation and expression of the cDNA,
RT comparison with human ARSB, and gene localization to feline
RT chromosome A1.";
RL Genomics 14:403-411(1992).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-
CC acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and
CC dermatan sulfate.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- DISEASE: DEFECTS IN ARSB ARE THE CAUSE OF MAROTEAUX-LAMY
CC SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).
CC MPS-VI HAS BEEN DESCRIBED IN SIAMESE CATS.
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S48472; AAB23941.1; -
DR PIR: A44475; A44475.
DR HSP: P15848; IfSU.
DR InterPro: IPR000917; Sulfatase.
DR Pfam: PF00884; Sulfatase; 1.
DR PROSITE: PS00523; SULFATASE.1; 1.
DR PROSITE: PS00149; SULFATASE.2; 1.
KW Hydrolase; Signal; Glycoprotein; Lysosome; Mucopolysaccharidosis.
FT CHAIN 1 41
FT SIGNAL 42 535
FT ACT_SITE 149 149
FT MOD_RES 93 93
FT FT 119 523
FT DISULFID 123 157
FT DISULFID 183 194
FT DISULFID 407 449
FT CARBOHYD 190 190
FT CARBOHYD 281 281
FT CARBOHYD 293 293
FT CARBOHYD 428 428
FT CARBOHYD 460 460
FT SEQUENCE 535 AA; 59753 MW; 43A527886A9983C4 CRC64;

Query Match 1.8%; Score 10; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 THWVGKWHLG 180
DB 142 THWVGKWHLG 151
|||||
```

```
RESULT 5
PUR6_SVNV3
ID PUR6_SVNV3 STANDARD; PRT; 176 AA.
AC Q55498;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoribosylaminoimidazole carboxylase catalytic subunit
DE (EC 4.1.1.21) (AIR carboxylase) (AIRC).
GN PURE OR SL0901.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- FUNCTION: THIS SUBUNIT CAN ALONE TRANSFORM AIR TO CAIR, BUT IN
CC ASSOCIATION WITH PURK, WHICH POSSESSES AN ATPASE ACTIVITY, AN
CC ENZYME COMPLEX IS PRODUCED WHICH IS CAPABLE OF CONVERTING AIR TO
CC CAIR EFFICIENTLY UNDER PHYSIOLOGICAL CONDITION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-5-amino-4-
CC imidazolecarboxylate - 1-(5-phosphoribosyl)-5-aminoimidazole +
CC CO(2).
CC -1- PATHWAY: SIXTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BACTERIAL PURE, ALSO TO AIR CARBOXYLASE FROM
CC FUNGI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D64006; BAA10848.1; -
DR HSP: P09028; IQC2.
DR InterPro: IPR000031; AIR_carboxyl.
DR Pfam: PF00731; AIRC; 1.
KW Purine biosynthesis; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 176 AA; 18603 MW; A1F87A630F18F8E4 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 QILASHNP 255
DB 131 QILASHNP 138
|||||

RESULT 6
ARS_HEMPU
ID ARS_HEMPU STANDARD; PRT; 551 AA.
AC P14000;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Arylsulfatase precursor (EC 3.1.6.1) (Aryl-sulfate sulphohydrolase)
DE (ARS).
OS Hemientrotus pulcherrimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Hemientrotus.
OX NCBI_TaxID=7650;
```


RA Churcher C., Coster F., Davis K., Davis R.W., Dietrich F.S.,
 RA Delius H., DiPaolo T., Dubois E., Dusterhoft A., Duncan M., Floeth M.,
 RA Fortin N., Friesen J.D., Fritz C., Goffeau A., Hall J., Hebling U.,
 RA Heumann K., Hilbert H., Hillier L., Hunnicke-Smith S., Hyman R., D.,
 RA Johnston M., Kalman S., Klein K., Komp C., Kurd O., Lashkari D.,
 RA Lew H., Lin A., Lin D., Louis E.J., Marathe R., Messenguy F.,
 RA Mewes H.W., Mirtipati S., Moestl D., Muller-Auer S., Namath A.,
 RA Newtlich U., Oefner P., Pearson D., Petel F.X., Pohl T.M.,
 RA Purnelle D., Schafer M., Scharfe M., Scherens B., Schramm S.,
 RA Schroeder M., Seicu A.M., Tettelin H., Urrestarazu L.A., Ushinsky S.,
 RA Vierendeels F., Viessers S., Voss H., Walsh S.V., Wambutt R., Wang Y.,
 RA Wedler E., Wedler H., Winnett E., Zhong W.W., Zollner A., Vo D.H.,
 RA Hani J.,
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";
 RL Nature 387:103-105(1997).
 RP [2]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE-21062923; PubMed-11112701;
 RA Park J.-H., Jensen B.C., Kifer C.T., Parsons M.;
 RT "A novel nucleolar G-protein conserved in eukaryotes.";
 RL J. Cell Sci. 114:173-185(2001).
 CC -|- SUBCELLULAR LOCATION: Nuclear; nucleolar.
 CC -|- SIMILARITY: BELONGS TO THE GTP1 / OBG FAMILY. NOG SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U43281; AAB68206.1; -;
 DR EMBL; U09176; AAA18614.1; -;
 DR EMBL; U33007; AAB64875.1; -;
 DR PIR; S48538; S48538.
 DR TRANSPAC; T02153; -;
 DR SGD; S0002851; SSN2.
 KW Nuclear protein.
 FT DOMAIN 526 529 POLY-ASN.
 FT DOMAIN 526 664 POLY-GLU.
 FT DOMAIN 813 816 POLY-SPR.
 FT DOMAIN 1005 1008 POLY-LEU.
 FT DOMAIN 1121 1136 POLY-GLN.
 FT CONFLICT 38 38 E -> E (IN REF. 2).
 FT CONFLICT 812 812 E -> V (IN REF. 2).
 FT CONFLICT 859 859 T -> S (IN REF. 2).
 FT CONFLICT 877 878 VK -> GE (IN REF. 2).
 FT CONFLICT 887 887 T -> P (IN REF. 2).
 FT CONFLICT 1284 1284 Y -> S (IN REF. 2).
 SQ SEQUENCE 647 AA; 74409 MW; 640324779AB4D716 CRC64;

 Query Match 1.5%; Score 8; DB 1; Length 647;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 48 QALEEEEE 55
 DB 454 QALEEEEE 461
 |||||

 RESULT 9
 SRB9_YEAST
 ID SRB9_YEAST STANDARD; PRT; 1420 AA.
 AC P38931;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DE 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Suppressor of RNA polymerase B SRB9 (SCAL protein).
 GN SRB9 OR SCAL OR SSN2 OR YDR443C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE-95293223; PubMed-7774808;
 RA Hengartner C.J., Thompson C.M., Zhang J., Chao D.M., Liao S.-M.,
 RA Koleske A.J., Okamura S., Young R.A.;
 RT "Association of an activator with an RNA polymerase II holoenzyme.";
 RL Genes Dev. 9:897-910(1995).
 RN [2]
 RP SEQUENCE FROM N.A.

RA Yuryev A., Corden J.L.;
 RA Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A.
 RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,
 RA Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
 RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,
 RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
 RA Winant A., Yelton M., Botstein D., Davis R.W.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 CC -|- FUNCTION: COMPONENT OF THE RNA POLYMERASE II HOLOENZYME AND THE
 CC MEDIATOR OF ACTIVATION SUBCOMPLEX.
 CC -|- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U23812; AAA91316.1; -;
 DR EMBL; U09176; AAA18614.1; -;
 DR EMBL; U33007; AAB64875.1; -;
 DR PIR; S48538; S48538.
 DR TRANSPAC; T02153; -;
 DR SGD; S0002851; SSN2.
 KW Nuclear protein.
 FT DOMAIN 526 529 POLY-ASN.
 FT DOMAIN 526 664 POLY-GLU.
 FT DOMAIN 813 816 POLY-SPR.
 FT DOMAIN 1005 1008 POLY-LEU.
 FT DOMAIN 1121 1136 POLY-GLN.
 FT CONFLICT 38 38 E -> E (IN REF. 2).
 FT CONFLICT 812 812 E -> V (IN REF. 2).
 FT CONFLICT 859 859 T -> S (IN REF. 2).
 FT CONFLICT 877 878 VK -> GE (IN REF. 2).
 FT CONFLICT 887 887 T -> P (IN REF. 2).
 FT CONFLICT 1284 1284 Y -> S (IN REF. 2).
 SQ SEQUENCE 1420 AA; 160000 MW; 7F6CF4BBE0FAC918 CRC64;

 Query Match 1.5%; Score 8; DB 1; Length 1420;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 153 LPDNLATL 160
 DB 1033 LPDNLATL 1040
 |||||

 RESULT 10
 CBIL_SALTY
 ID CBIL_SALTY STANDARD; PRT; 237 AA.
 AC Q05593;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Precorrin-2 C20-methyltransferase (EC 2.1.1.130) (S-adenosyl-L-
 DE methionine--precorrin-2 methyltransferase) (SP2MT).
 GN CBIL OR STM2024.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.
 RC STRAIN-L72;
 RX MEDLINE-93273696; PubMed=8501034;
 RA Roth J.R., Lawrence J.G., Rubinfeld M., Kieffer-Higgins S.,
 RA Church G.M.;


```

RT *Characterization of the cobalamin (vitamin B12) biosynthetic genes
RT of Salmonella typhimurium.
RL J. Bacteriol. 175:3303-3316(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA *Complete genome sequence of Salmonella enterica serovar Typhimurium
RT J2;
RL Nature 413:852-856(2001).
RC -1- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC PRECORRIN-3A.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 -> S-
CC adenosyl-L-homocysteine + precorrin-3A.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC AND CBIL/COBI.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L12006; AAA27263.1; -.
CC EMBL; AE008789; AAL20928.1; -.
CC StyGene; SG10044; cbil.
CC InterPro; IPR000878; Corrin_porph_methyltrnf.
CC Pfam; PF00590; TP_methylase; 1.
CC PROSITE; PS00839; SUMT.1; 1.
CC PROSITE; PS00840; SUMT.2; 1.
CC Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
CC Methyltransferase; Complete proteome.
CC CONFLICT 188 188 Q -> T (IN REF. 1).
CC SEQUENCE 237 AA; 25806 MW; 3D7215BDC67D2067 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 237;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ALLAAQAG 63
DB 184 ALLAAQAG 190
|||||||

RESULT 11
VGLG_HSV11
ID VGLG_HSV11 STANDARD; PRT; 238 AA.
AC P06484;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Glycoprotein G.
GN GG OR US4.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85160822; PubMed=2984429;
RA McGeoch D.J., Dolan A., Donald S., Rixon F.J.;
RT *Sequence determination and genetic content of the short unique

```

```

RT region in the genome of herpes simplex virus type 1.
RL J. Mol. Biol. 181:1-13(1985).
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC 2: GH, GB, GC, GG, GD, GI, AND GE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00036; AAA96684.1; -.
CC EMBL; X14112; CAA32281.1; -.
CC EMBL; X02138; CAA26058.1; -.
CC PIR; A05239; QQBE74.
CC Glycoprotein. 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 238 AA; 25238 MW; F7FDC2867E834B92 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 EEEEEGA 57
DB 81 EEEEEGA 87
|||||||

RESULT 12
PSA3_ACACA
ID PSA3_ACACA STANDARD; PRT; 252 AA.
AC P90513;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Proteasome subunit alpha type 3 (EC 3.4.25.1) (Fragment).
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RA Xu P., Zot H.G.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
CC WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,
CC PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR
CC SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC
CC ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
CC specificity.
CC -1- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal
CC proteolytic pathway.
CC -1- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
CC SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A; ALSO KNOWN AS THE
CC PROTEASOME A-TYPE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U85398; AAB41645.1; -.
CC HSSP; P25156; IPMA.

```

```

DR MEROPS; T01.977; -.
DR InterPro; IPR001353; Proteasome.
DR InterPro; IPR000426; Proteasome_A.
DR Pfam; PF00227; proteasome; 1.
DR PROSITE; PS00388; PROTEASOME_A; 1.
KW Proteasome; Hydrolase; Protease.
FT NON_TER 1
SQ SEQUENCE 252 AA; 27728 MW; 71898276B50C2E79 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 QALEEE 54
DB 239 QALEEE 245
|||||

RESULT 13
PANC_NEIMB STANDARD; PRT; 278 AA.
AC P57036;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Pantoate-beta-alanine ligase (EC 6.3.2.1) (Pantothenate synthetase)
DE (Pantoate activating enzyme).
GN PANC OR NMB0871.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citron H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815(2000).
CC -1- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP +
CC diphosphate + (R)-pantothenate.
CC -1- PATHWAY: Pantothenate biosynthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE PANTOTHENATE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002439; AAP41282.1; -.
DR TIGR; NMB0871; -.
DR InterPro; IPR003721; Pantoate_ligase.
DR Pfam; PF02569; Pantoate_ligase; 1.
KW Pantothenate biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 278 AA; 31080 MW; 5E71D92FB0E6F6B3 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 DKLAEG 110

```

```

DB 76 DKLAEG 82
|||||

RESULT 14
HSLO_SYNY3 STANDARD; PRT; 302 AA.
AC P73910;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 33 kDa chaperonin (Heat shock protein 33 homolog) (HSP33).
GN HSLO OR SLL1988.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Suglura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: REDOX REGULATED MOLECULAR CHAPERONE. PROTECTS BOTH
CC THERMALLY UNFOLDING AND OXIDATIVELY DAMAGED PROTEINS FROM
CC IRREVERSIBLE AGGREGATION. PLAYS AN IMPORTANT ROLE IN THE BACTERIAL
CC DEFENSE SYSTEM TOWARD OXIDATIVE STRESS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- PTM: UNDER OXIDIZING CONDITIONS TWO DISULFIDE BONDS ARE FORMED
CC INVOLVING THE REACTIVE CYSTEINES. UNDER REDUCING CONDITIONS ZINC
CC IS BOUND TO THE REACTIVE CYSTEINES AND THE PROTEIN IS INACTIVE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HSP33 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90910; BAA1974.1; -.
DR InterPro; IPR000397; Hsp33.
DR Pfam; PF01430; HSP33; 1.
DR ProDom; PD014025; Hsp33; 1.
KW Chaperone; Redox-active center; Zinc; Complete proteome.
FT DISULFID 240 242 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 273 276 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 302 AA; 32205 MW; 0EB3448B06E7AAA CRC64;

Query Match 1.3%; Score 7; DB 1; Length 302;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 351 GGIRAVG 357
DB 13 GGIRAVG 19
|||||

RESULT 15
TRUB_BACHD STANDARD; PRT; 304 AA.
ID TRUB_BACHD
AC Q9KA80;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55
DE synthase) (Psi55 synthase) (pseudouridylylate synthase) (Uracil
DE hydrolase).
GN TRUB OR BH2410.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125/JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF001515; BAB06129.1; -.
DR InterPro: IPR002501; TruB_N.
DR Pfam; PF01509; TruB_N; 1.
KW Lyase; tRNA processing; Complete proteome.
SQ SEQUENCE 304 AA; 34054 MW; 2DE397CE20C67946 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 EGGALLA 60
|
Db 275 EGGALLA 281

Search completed: July 23, 2002, 19:10:27
Job time: 152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:06:45 ; Search time 28.8 seconds
(without alignments)
3303.723 Million cell updates/sec

Title: US-09-495-823-7
Perfect score: 550
Sequence: 1 MAPRCAGHPPPSPQACV.....VPRYPKDRSNRLNGGV 550

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTEMBL19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	11.5	169	4 Q9H5D8	Q9h5d8 homo sapien
2	20	3.6	266	5 Q9NJU7	Q9nJU7 helix pomat
3	12	2.2	503	5 Q9NJU8	Q9nJU8 helix pomat
4	8	1.5	217	5 Q9VQ95	Q9vQ95 drosophila
5	8	1.5	330	16 Q9HYR2	Q9hyr2 pseudomonas
6	8	1.5	381	16 Q92EX3	Q92ex3 listeria in
7	8	1.5	452	5 Q18924	Q18924 caenorhabdi
8	8	1.5	433	16 Q9CKE0	Q9cke0 pasteurella
9	8	1.5	542	5 Q9V6F2	Q9v6f2 drosophila
10	8	1.5	551	5 Q25384	Q25384 hemiceirotr
11	8	1.5	559	5 Q16138	Q16138 heliocidari
12	8	1.5	579	5 Q9VVM4	Q9vVM4 drosophila
13	8	1.5	898	2 Q93D92	Q93d92 streptococ
14	8	1.5	936	5 Q9VVK9	Q9vVK9 drosophila
15	8	1.5	1002	10 Q9AXB1	Q9axb1 oryza sativ
16	8	1.5	1216	16 Q92DI9	Q92di9 listeria in

17	7	1.3	24	4	Q96I53	Q96I53 homo sapien
18	7	1.3	27	15	O11847	O11847 human immun
19	7	1.3	48	5	Q9VWS6	Q9vWS6 drosophila
20	7	1.3	79	12	Q68100	Q68100 human cytom
21	7	1.3	105	8	Q9B267	Q9b267 pachydaetyl
22	7	1.3	114	16	O66649	O66649 aquifex aeo
23	7	1.3	123	16	Q9K8T2	Q9k8t2 bacillus ha
24	7	1.3	123	17	Q9Y992	Q9y992 aeropyrum p
25	7	1.3	126	12	O65435	O65435 beet curly
26	7	1.3	128	8	O78764	O78764 pyxis plani
27	7	1.3	128	12	Q68084	Q68084 human cytom
28	7	1.3	141	16	Q98CC0	Q98cc0 rhizobium l
29	7	1.3	143	4	Q96K66	Q96k66 homo sapien
30	7	1.3	158	12	O91TP8	Q91tp8 tupaia herp
31	7	1.3	159	16	O92DS6	Q92ds6 listeria in
32	7	1.3	163	5	Q9NB86	Q9nb86 agrotis ips
33	7	1.3	164	2	P71285	P71285 escherichia
34	7	1.3	169	2	O50216	O50216 pseudomonas
35	7	1.3	169	2	O33496	O33496 pseudomonas
36	7	1.3	169	2	O06836	O06836 pseudomonas
37	7	1.3	172	12	O91D82	Q91d82 human cytom
38	7	1.3	172	15	O9ORM7	Q9orm7 human immun
39	7	1.3	175	10	Q9FM63	Q9fm63 arabidopsis
40	7	1.3	179	10	Q94LB7	Q94lb7 oryza sativ
41	7	1.3	180	2	Q01092	Q01092 serratia ma
42	7	1.3	194	5	Q9U933	Q9u933 paramecium
43	7	1.3	198	3	P87027	P87027 schizosacch
44	7	1.3	199	3	O14312	O14312 schizosacch
45	7	1.3	199	4	Q9BUI1	Q9bull homo sapien

ALIGNMENTS

RESULT 1

Q9H5D8 ID Q9H5D8 PRELIMINARY; PRT; 169 AA.
AC Q9H5D8;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE CDNA: FLJ23548 FIS, CLONE LMG08487.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Kawakami T., Nozuchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hiraio M., Ohmori Y., Ota T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027201; BAB15689.1; -
DR HSSP; P15848; IFSU.
SQ SEQUENCE 169 AA; 18905 MW; 4583C92FFFC3380 CRC64;

Query Match 11.5%; Score 63; DB 4; Length 169;
Best Local Similarity 100.0%; Pred. No. 4.7e-57;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 488 STGKSVLNFNITADPYERVVLSNRYPGIVKKLLRRLSQFNKTAVPVRYPPKDRSNRPNLN 547

Db 58 STGKSVLNFNITADPYERVVLSNRYPGIVKKLLRRLSQFNKTAVPVRYPPKDRSNRPNLN 117

Qy 548 GGV 550

Db 118 GGV 120

RESULT 2
ID Q9NJU7 PRELIMINARY; PRT; 266 AA.
AC Q9NJU7; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SULFATASE 2 (FRAGMENT).
GN SULF2.
OS Helix pomatia (Roman snail) (Edible snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Helicaceae; Helicidae; Helix.
OX NCBI_TaxID=6536;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20233298; PubMed=10772344;
RA Wittstock U., Fischer M., Svendsen I., Halkier B.A.;
RT "Cloning and characterization of two cDNAs encoding sulfatases in the
RL Roman snail, Helix pomatia.";
RL IUBMB Life 49:71-76(2000).
DR EMBL; AF109925; AAF30403.1; -.
DR HSSP; AF15848; IFSU.
DR InterPro; IPR000917; Sulfatase.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
FT NON_TER 266 266
SQ SEQUENCE 266 AA; 30391 MW; 2F877BB1737A9C49 CRC64;

Query Match 3.6%; Score 20; DB 5; Length 266;
Best Local Similarity 100.0%; Pred. No. 4.6e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 103 LDKLAAGVKLENYVOPIC 122
Db 55 LDKLAAGVKLENYVOPIC 74
|||||

RESULT 3
ID Q9NJU8 PRELIMINARY; PRT; 503 AA.
AC Q9NJU8; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SULFATASE 1.
GN SULF1.
OS Helix pomatia (Roman snail) (Edible snail).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC Helicaceae; Helicidae; Helix.
OX NCBI_TaxID=6536;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20233298; PubMed=10772344;
RA Wittstock U., Fischer M., Svendsen I., Halkier B.A.;
RT "Cloning and characterization of two cDNAs encoding sulfatases in the
RL Roman snail, Helix pomatia.";
RL IUBMB Life 49:71-76(2000).
DR EMBL; AF109924; AAF30402.1; -.
DR HSSP; AF15848; IFSU.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
SQ SEQUENCE 503 AA; 55821 MW; AC45334BA694413D CRC64;

Query Match 2.2%; Score 12; DB 5; Length 503;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 113 LENYVQPICTP 124
Db |||||||

Db 71 LENYVQPICTP 82

RESULT 4
ID Q9VQ95 PRELIMINARY; PRT; 217 AA.
AC Q9VQ95; 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CGI7006 PROTEIN.
GN CGI7006.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003583; AAF51282.1; -.
DR FlyBase; FBgn0031404; CGI7006.
DR InterPro; IPR001604; Endonuclease.
DR SMART; SM00477; NUC; 1.
SQ SEQUENCE 217 AA; 24439 MW; ECABDBF82A5533C2 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 217;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 ERVLSNR 511
Db |||||||
35 ERVLSNR 42

```

RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596164; CAC95568.1; -.
DR ListList; LINO0335; -.
KW Complete proteome.
SQ SEQUENCE 381 AA; 43625 MW; D74B4FEE1C1A720C CRC64;

Query Match 1.5%; Score 8; DB 16; Length 381;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 GFYNNII 322
DB 181 GFYNNII 188
|||||||

RESULT 7
Q18924 PRELIMINARY; PRT; 452 AA.
ID Q18924
AC Q18924;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 51.7 KDA PROTEIN.
GN D1014.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du 2., Leimbac D.;
RT "The sequence of C. elegans cosmid D1014.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53180; AAA96290.1; -.
DR HSSP; P15289; LAUK.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 452 AA; 51685 MW; 14999D3C1F7891D2 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 452;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180
DB 134 MVGKWHLG 141
|||||||

RESULT 8
Q9CKEO PRELIMINARY; PRT; 453 AA.
ID Q9CKEO
AC Q9CKEO;

RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596164; CAC95568.1; -.
DR ListList; LINO0335; -.
KW Complete proteome.
SQ SEQUENCE 381 AA; 43625 MW; D74B4FEE1C1A720C CRC64;

Query Match 1.5%; Score 8; DB 16; Length 330;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 EEGGALLA 60
DB 171 EEGGALLA 178
|||||||

RESULT 6
Q92EX3 PRELIMINARY; PRT; 381 AA.
ID Q92EX3
AC Q92EX3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LINO335 PROTEIN.
GN LINO335.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
Gautier L., Goebel W., Gomez-Lopez N., Hain J., Hauf J., Jackson D.,
Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
Rammel B., Rose M., Schlueder T., Simoes N., Tierrez A.,
Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;

Query Match 1.5%; Score 8; DB 16; Length 330;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 EEGGALLA 60
DB 171 EEGGALLA 178
|||||||

RESULT 6
Q92EX3 PRELIMINARY; PRT; 381 AA.
ID Q92EX3
AC Q92EX3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LINO335 PROTEIN.
GN LINO335.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
Gautier L., Goebel W., Gomez-Lopez N., Hain J., Hauf J., Jackson D.,
Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
Rammel B., Rose M., Schlueder T., Simoes N., Tierrez A.,
Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;

```

DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN PM1682.
 GN PM1682.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RY May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006204; AK03766.1; -;
 DR InterPro: IPR000566; Lipocin_cyFABP.
 DR Pfam: PF00884; Sulfatase; 1.
 DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 453 AA; 51449 MW; 83B91532C6252B0E CRC64;

Query Match 1.5%; Score 8; DB 16; Length 453;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 TPTLDKLA 107
 |||||
 Db 29 TPTLDKLA 36

RESULT 9

ID Q3V6F2 PRELIMINARY; PRT; 542 AA.
 AC Q3V6F2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE C68646 PROTEIN.
 GN C68646.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rainer K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003821; AAF58475.1; -;
 DR HSSP: P15848; IFSU.
 DR FlyBase; FBgn0033763; CG8646.
 DR InterPro: IPR000917; Sulfatase.
 DR Pfam: PF00884; Sulfatase; 1.
 DR PROSITE: PS00523; SULFATASE_1; 1.
 DR PROSITE: PS00149; SULFATASE_2; 1.
 SQ SEQUENCE 542 AA; 60278 MW; EACD9C14DD2B3204 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 542;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 PICTPSRS 127
 |||||
 Db 71 PICTPSRS 78

RESULT 10

ID Q25384 PRELIMINARY; PRT; 551 AA.
 AC Q25384;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE ARYL SULFATASE.
 GN ARS.
 OS Hemientrotus pulcherrimus (Sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Hemientrotus.
 OX NCBI_TaxID=7650;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SPERM;
 RA Yamada K.;
 RL Submitted (SEP-1989) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE OF 1-105 FROM N.A.
 RC TISSUE=SPERM;
 RX MEDLINE=90092130; PubMed=2598936;
 RA Yamada K., Akasaka K., Shimada H.;
 RT "Structure of sea-urchin arylsulfatase gene.";
 RL Eur. J. Biochem. 186:405-410(1989).
 [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SPERM;
 RA Shimada H.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SPERM;
 RA Akasaka K., Sakamoto N., Yamamoto T., Morokuma J., Fujikawa N.,
 RA Takata K., Eguchi S., Shimada H.;
 RT "Corrected structure of the 5' flanking region of arylsulfatase gene of

RT the sea urchin, Hemicentrotus pulcherrimus.*;
 RL Dev. Growth Differ. 36:633-636(1994).
 DR EMBL: X16679; CAA34667.1; .
 DR EMBL: X16680; CAA34667.1; JOINED.
 DR EMBL: X16681; CAA34667.1; JOINED.
 DR EMBL: X16682; CAA34667.1; JOINED.
 DR EMBL: X16683; CAA34667.1; JOINED.
 DR EMBL: X16684; CAA34667.1; JOINED.
 DR HSSP: P15289; IAUk.
 DR InterPro: IPR000917; Sulfatase.
 DR Pfam: PF00884; Sulfatase; 1.
 DR PROSITE; PS00523; SULFATASE_1; 1.
 DR PROSITE; PS00149; SULFATASE_2; 1.
 SQ SEQUENCE 551 AA; 60943 MW; 544315A4880E12C1 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 551;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180
 |||||

Db 153 MVGKWHLG 160

RESULT 11

O16138 PRELIMINARY; PRT; 559 AA.
 AC O16138;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE ARYL-SULFATASE.
 GN ARS.
 OS Helicoidaris erythrogramma (Sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinometridae;
 OC Helicoidaris.
 OX NCBI_TaxID=7634;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98298333; PubMed=9634485;
 RA Haag E.S., Raff R.A.;

RT "Isolation and characterization of three mRNAs enriched in embryos of
 the direct-developing sea urchin Helicoidaris erythrogramma: evolution
 of larval ectoderm.";
 RT Dev. Genes Evol. 208:188-204(1998).
 RL EMBL: AF013158; AAC27821.1; .
 DR HSSP: P15289; IAUk.
 DR InterPro: IPR000917; Sulfatase.
 DR Pfam: PF00884; Sulfatase; 1.
 DR PROSITE; PS00523; SULFATASE_1; 1.
 DR PROSITE; PS00149; SULFATASE_2; 1.
 SQ SEQUENCE 559 AA; 61915 MW; 13DE7DCEBD05752 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 559;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180
 |||||

Db 159 MVGKWHLG 166

RESULT 12

Q9VVM4 PRELIMINARY; PRT; 579 AA.
 AC Q9VVM4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG7402 PROTEIN.

GN CG7402.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M., Pan Z.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Fang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003522; AAF49287.1; .
 DR HSSP: P15848; 1FSU.
 DR FlyBase; FBgn0036768; CG7402.
 DR InterPro: IPR000917; Sulfatase.
 DR Pfam: PF00884; Sulfatase; 1.
 DR PROSITE; PS00149; SULFATASE_2; 1.
 SQ SEQUENCE 579 AA; 65053 MW; EBF2F4F13D759298 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 579;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 VGRKHLGF 181
 |||||

Db 127 VGRKHLGF 134

RESULT 13

Q93D92 PRELIMINARY; PRT; 898 AA.
 ID Q93D92
 AC Q93D92;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

```
DE PSAB.
GN PSAB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=L711;
RA Tao L., Tanzer J.M.;
RT "Novel sucrose-dependent adhesion cofactors (sdc) in Streptococcus
RT mutans.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF397166; AAL04089.1; -.
SQ SEQUENCE 898 AA; 101903 MW; A2D9CA6F9357AF11 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 898;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 TFFGSLLC 202
Db 440 TFFGSLLG 447
|||||||

RESULT 14
Q9VVK9 PRELIMINARY; PRT; 996 AA.
AC Q9VVK9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG5584 PROTEIN.
GN CG5584.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
```

```
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003522; AAP49302.1; -.
DR HSP; F15848; IfSU.
DR FlyBase; FBgn0036755; CG5584.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00149; SULFATASE-2; 1.
SQ SEQUENCE 996 AA; 113514 MW; C92AA378D082FB3 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 996;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 VGKWLGF 181
Db 126 VGKWLGF 133
|||||||

RESULT 15
Q9AXB1 PRELIMINARY; PRT; 1002 AA.
ID Q9AXB1;
AC Q9AXB1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE P0501G01.6 PROTEIN.
GN P0501G01.6.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0501G01.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002819; BAB21077.1; -.
DR InterPro; IPR001871; bZIP.
DR InterPro; IPR001296; Glycos.transf_1.
DR Pfam; PF00534; Glycos.transf_1; 1.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ SEQUENCE 1002 AA; 113136 MW; F459312DD89CC5EF CRC64;

Query Match 1.5%; Score 8; DB 10; Length 1002;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GHPPPPSP 15
Db 24 GHPPPPSP 31
|||||||
```

Search completed: July 23, 2002, 19:09:59
Job time: 194 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:02:54 ; Search time 38.31 Seconds
(without alignments)
1594.638 Million cell updates/sec

Title: US-09-495-823-7
Perfect score: 550
Sequence: 1 MAPRCAGHPPPSPQACV.....VPRYPKDRSPRLNGGV 550

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0 747574

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	385	70.0	599	22 AAB85483	Human 23553 sulfat
2	289	52.5	289	22 AAU14421	Human novel protei
3	196	35.6	22	AAU14185	Human novel protei
4	181	32.9	515	20 AAY41701	Human PRO708 prote
5	181	32.9	515	21 AAB44257	Human PRO708 (UNQ3
6	181	32.9	515	22 AAU29061	Human PRO polypept
7	179	32.5	187	22 AAM25415	Human protein sequ
8	109	19.8	128	22 AAU23335	Novel human enzyme
9	19	3.5	569	22 AAB85482	Human 25278 sulfat
10	10	1.8	533	22 AAB51184	Human sulfatase pr
11	8	1.5	217	22 ABB66123	Drosophila melanog

12	8	1.5	330	22 AAU36321	Pseudomonas aerugi
13	8	1.5	542	22 ABB63962	Drosophila melanog
14	8	1.5	551	22 AAB51186	Sea urchin sulfata
15	8	1.5	579	22 ABB63246	Drosophila melanog
16	8	1.5	996	22 ABB61654	Drosophila melanog
17	8	1.5	1420	18 AAW13826	Yeast transcrip tio
18	7	1.3	15	22 AAB86556	Human cytomegalovi
19	7	1.3	15	22 AAB86556	Human cytomegalovi
20	7	1.3	15	22 AAB86557	Human cytomegalovi
21	7	1.3	16	22 AAB86557	Human cytomegalovi
22	7	1.3	16	22 AAB46626	HIV-1 Tat peptide
23	7	1.3	19	21 AAB10389	HPV E6 protein inh
24	7	1.3	28	14 AAR39867	E2 peptide E2-15,
25	7	1.3	34	22 AAO12394	Human polypeptide
26	7	1.3	45	22 AAM23940	Human EST encoded
27	7	1.3	48	22 ABB70089	Drosophila melanog
28	7	1.3	51	22 AAM83802	Human immune/haema
29	7	1.3	55	22 AAU22229	Human cardiovascular
30	7	1.3	62	22 ABB27456	Human peptide #107
31	7	1.3	62	22 ABB32605	Peptide #111 encod
32	7	1.3	62	22 ABB18101	Protein #100 encod
33	7	1.3	62	22 AAM53434	Human brain expres
34	7	1.3	62	22 AAM13674	Human bone marrow
35	7	1.3	62	22 AAM65812	Peptide #108 encod
36	7	1.3	62	22 AAM26073	Peptide #110 encod
37	7	1.3	75	22 AAM01424	Peptide #106 encod
38	7	1.3	79	22 AAU29568	Novel human secret
39	7	1.3	85	22 AAG01102	Novel human diagno
40	7	1.3	86	22 AAU29804	Novel human secret
41	7	1.3	105	22 AAB46684	HIV-1 Tat peptide
42	7	1.3	105	22 AAO06154	Human polypeptide
43	7	1.3	110	22 AAU31767	Novel human secret
44	7	1.3	112	22 AAG06329	Novel human diagno
45	7	1.3	113	22 AAG01780	Human polypeptide
			115	21 AAG57479	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AA885483	
ID	AAB85483 standard; Protein; 599 AA.
XX	
AC	AAB85483;
XX	
DT	25-SEP-2001 (first entry)
XX	
DE	Human 23553 sulfatase polypeptide.
XX	
SW	Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;
KW	antiinflammatory; vasotropic; antitumor; gene therapy; human.
XX	
OS	Homo sapiens.
XX	
PN	WO200155411-A2.
XX	
PD	02-AUG-2001.
XX	
PF	31-JAN-2001; 2001WO-US03266.
XX	
PR	31-JAN-2000; 2000US-0495823.
XX	
PI	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
XX	
DR	WPI; 2001-476214/51.
XX	
DR	N-PSDB; AAB46863, AAB46867.
XX	
PT	Novel human sulfatase polypeptides useful for treating and diagnosing
PT	sulfatase-related disorders such as cerebrovascular diseases, acute
XX	meningitis, multiple sclerosis, degenerative diseases and tumor -
XX	

PS Claim 9; Fig 15; 180pp; English.

XX The invention provides 22438, 23553, 25278 or 26212 human sulfatase
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides
CC are useful for treating disorders involving the brain such as
CC cerebrovascular diseases, infections such as acute meningitis,
CC demyelinating diseases including multiple sclerosis, degenerative
CC diseases affecting the cerebral cortex including Alzheimer's disease
CC and Pick disease, spinocerebellar degenerations including spinocerebellar
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,
CC degenerative diseases affecting motor neurons including amyotrophic
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,
CC toxic and acquired metabolic diseases, including vitamin deficiencies,
CC and neurocutaneous syndromes (phakamatoses) including neurofibromatosis.
CC The present sequence represents a human 23553 sulfatase polypeptide.

XX Sequence 599 AA;

Query Match 70.0%; Score 385; DB 22; Length 599;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 485; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAPRGAGHPPPSPQACVCPGKMLAGFWILCLTYGLSWGQALEEEEGALLA 60
DB 1 maprgcaghppspqacvcpqkmlangalagfwilclltygylswgaleeeegalla 60
QY 61 QAGCKLEPSTTSQPHLIILADDDQFRDVGHGSEIKTPTLDKLAAGVKLENYVQP 120
DB 61 qagcklepsttsqphliiladddqfrdvghgseiktptldklaagvklennyvqp 120
QY 121 ICTPSRQFTGKYQIHTGLQHSIIRTPQNCPLDNLATPQKLKEGYSTHMGKWHLG 180
DB 121 ictpsrqftgkyqihgtlqhsirtpqncpldnlatlqklkegysthmgkwhlg 180
QY 181 FYRKECMPTRRGFTFGSLGSDGYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQ 240
DB 181 fyrkecmptrrgftfgslgsdgythykcdspgcmgydlyendnaawdydngiystq 240
QY 241 MYTORVOQILASHNPTKPIFLYTAQVHSPLOQAPGRYFEHYRSIINRRYAAMLSCL 300
DB 241 mytrvqqlashnptkpfilytaqvahsplqapgrfyehyrsiinrryaaamlsc 300
QY 301 DEALNNVTALKTYGFYNNSSIIYSSDNGGQPTAGGSNWPLRSGKTYWEGGIRAVGFVH 360
DB 301 dealnvtlalktygfynssiiyssdnggqptaggsnwplrgsktyweggiravgf 360
QY 361 SPLKNGKTVCKEVPVHTDWPYPTLISLAEGQIDEDIQLDGYDIWETISEGLRSPRVILH 420
DB 361 spllknkgtvckelvhitdwpypylislaegqidediqldgydiwetiseglrsprv 420
QY 421 NIDPIYTKAKNGSWAAGYGIWNTAIOAIRVOHWKLLTGNPGYSDWVPQPSFSLGNRW 480
DB 421 nidpiytkakngswaagygiwntaioairvohwklltgnpgysdwvppqpsfslgn 480
QY 481 HNERIT 486
DB 481 hnerit 486

RESULT 2

ID AAU14421
XX AAU14421 standard; Protein; 289 AA.

AC AAU14421;

XX 24-OCT-2001 (first entry)

XX Human novel protein #292.

XX Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnerable; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;

KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX tissue regeneration; immune disorder.

OS Homo sapiens.

PN W0200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-451939/48.

XX N-PSDB; AAS22726.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,

XX nervous system disorders, and for regenerating bone and cartilage -

XX Example 4; Page 809-810; 894pp; English.

CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/ elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence represents a protein of the invention.

XX Sequence 289 AA;

Query Match 52.5%; Score 289; DB 22; Length 289;
Best Local Similarity 100.0%; Pred. No. 2.4e-285;
Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 MLAMGALAGFWILCLTYGLSWGQALEEEEGALLAAGKLEPSTTSQPHLIIFILA 83

DB 1 mlamgalagfwilclltygylswgaleeeegallagakelepsttsqphlilila 60

QY 84 DDQCFRDVGHGSEIKTPTLDKLAAGVKLENYVQPICTPSRQFTGKYQIHTGLQHS 143

DB 61 ddqgfrdvghgseiktptldklaagvklennyvqpictpsrqftgkyqihgtlqhs 120

QY 144 IIRTPQNCPLDNLATPQKLKEGYSTHMGKWHLGFRYRKECMPTRRGFTFGSLG 203

DB 121 iirtpqncpldnlatlqklkegysthmgkwhlgfyrkecmptrrgftfgslg 180

QY 204 GDYVTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQMYTORVOQILASHNPTKPIFLYI 263

Db 181 gdyvthykcdspgncgydlyendaawdydngiydstqmytqrvcqilashnptkpfilyi 240
Qy 264 AYQAVHSPLOAPGRYFHEHRSIIINRRRYAAMLSCIDEAINNVTLALK 312
Db 241 ayqavhsplqapgrfyfhehrysiinihrrryaamiscideainnvtlalk 289
RESULT 3
AAU14185
ID AAU14185 standard; Protein; 366 AA.
XX
AC AAU14185;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel protein #56.
XX
KW Human; novel protein; Antianaemic; osteopathic; antinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI: 2001-451939/48.
DR N-PSDB: AAS22490.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Example 4; Page 557-558; 894pp; English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
CC The present sequence represents a protein of the invention.

XX
SQ Sequence 366 AA;
Query Match 35.6%; Score 196; DB 22; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.3e-190;
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 16 QACVCPGKMLAMGALAGFWILCLLTGYLSWGQALEEEEGALLAQAGEKLEPSTTSQ 75
Db 162 qacvcpgkmlamgalagfwilclltygylswgqaleeeegallaqageklepsttsq 221
Qy 76 PHLIFILADQGFDRVGVHSGSEIKTPTLDKLAABGVKLENYVQPICTPSRSQFITGKYQ 135
Db 222 phliladddqgfrdvgyhgseiktptldklaaegvklenyvqpictpsrsqfitgkyq 281
Qy 136 IHTGLQHSIIIRPTQNCPLDNLATLPQKLKEVGYSTHVMGKWHLGFYRKECMPTRRGDT 195
Db 282 ihtglqhsilrptqncpldnltpqklkevgysthvmgkwhlgfyркеcmptrrgdt 341
Qy 196 PFGSLGSGDYTHYK 211
Db 342 ffgslgsgdythyk 357
RESULT 4
AAV41701
ID AAV41701 standard; Protein; 515 AA.
XX
AC AAV41701;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO708 protein sequence.
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
XX
PN WO9946281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.

CC the proliferation or differentiation of chondrocyte cells. The PRO
CC proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
SQ Sequence 515 AA;

Query Match 32.9%; Score 181; DB 22; Length 515;
Best Local Similarity 100.0%; Pred. NO. 3.4e-175;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPRCAGHPPPPSPQACVCPGKMLAMCALAGFWILCLLTGYLSWGOALREEEGALLA 60
Db 1 mapr9caghp9ppspqacvcp9gkmlangatagfwilclltgy1swgqaleeeegalla 60
|||||
Qy 61 QAGEKLEPSTTSQPHLFIILADDQGFDRDVGYPHSGSEIKTPTLDKLAEGVKLENYYVQP 120
Db 61 qageklepsttsqphlfiiladdqgfrdvgyhgseikptldklaaegvklennyvqp 120
|||||
Qy 121 ICTPSRQFTGKQIHTGLQHSIRTPQNCPLDNPATLPQKLEKVGYSTHMYGKWHLG 180
Db 121 ictpsrqftgkqihgtglqhsirtpqncpldnpatlpqk1ekvgysthmvgkwhlg 180
|||||
Qy 181 F 181
Db 181 f 181

RESULT 7
AAU25415
ID AAU25415 standard; Protein; 187 AA.
XX
AC AAU25415;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:930.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnery; antilucer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.
XX
OS Homo sapiens.
XX
XX WO200153455-A2.
XX
XX 26-JUL-2001.
XX
XX 22-DEC-2000; 2000WO-US35017.
XX
XX 23-DEC-1999; 99US-0471275.
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX

DR WPI; 2001-457603/49.
DR N-PSDB; AAH99356.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
PS Claim 20; Page 200; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
CC antilucer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis of
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
SQ Sequence 187 AA;

Query Match 32.5%; Score 179; DB 22; Length 187;
Best Local Similarity 100.0%; Pred. NO. 1.5e-173;
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 YOIHTGLQHSIRTPQNCPLDNPATLPQKLEKVGYSTHMYGKWHLGFKYRKECMPTRRGF 193
Db 9 yqihgtglqhsirtpqncpldnpatlpqk1ekvgysthmvgkwhlgfyrkecmptrrgf 68
|||||
Qy 194 DTFEGSLGSGDYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQMYTORVOQILASH 253
Db 69 dtfgslgsgdytyhykcdspgcmgydlyendnaawdydnglystqmytrvqgillash 128
|||||
Qy 254 NPTKPIFLYIAQVHSPLOAPGRVFEHYRSIINIRRYAAMLSCLDENNNVTALK 312
Db 129 nptkpiflyiaqvahsp1qapgrfyehrsilnirryaam1scldena1nnvtalk 187
|||||

RESULT 8
AAU23335
ID AAU23335 standard; Protein; 128 AA.
XX
AC AAU23335;
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #421.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
XX WO200155301-A2.
XX
XX 02-AUG-2001.
XX

PF 17-JAN-2001; 2001WO-US01239.
XX
31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465566/50.
DR N-PSDB; AAS41205.
XX

PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases
XX
PS Claim 11; SEQ ID No 1331; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 128 AA;

Query Match 19.8%; Score 109; DB 22; Length 128;
Best Local Similarity 100.0%; Pred. No. 1.8e-102;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 90 DVGTHGSEIKPTLDKLAAGVCKLENYYVQPICTPSPQSOFITGKYQHTGQLQHSIIIRPTQ 149
Db 6 dvgyhgseiktptldklaagvcklenyyvqpictpsrsqfitgkyqhtgqlqhsilrptq 65
QY 150 PNCPLDNATLPQKLKEVGSTHMGVGHGLGFYRKECMPTRRGDTFFG 198
Db 66 pncpldnatlpqklkevgsthmvgvghlgfyrkcmptrrgidtffg 114

RESULT 9
AAB85482
ID AAB85482 standard; Protein; 569 AA.
XX
AC AAB85482;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human 25278 sulfatase polypeptide.
XX
KW Sulfatase; 25278 sulfatase; nootropic; neuroprotective; antibacterial;
KW antiinflammatory; vasotropic; antitumor; gene therapy; human.
XX
OS Homo sapiens.
XX
PN W0200155411-A2.
XX
PD 02-AUG-2001.
XX
PF 31-JAN-2001; 2001WO-US03266.
XX
PR 31-JAN-2000; 2000US-0495823.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
XX WPI; 2001-476214/51.
DR N-PSDB; AAH46862, AAH46866.
XX

PT Novel human sulfatase polypeptides useful for treating and diagnosing
PT sulfatase-related disorders such as cerebrovascular diseases, acute
PT meningitis, multiple sclerosis, degenerative diseases and tumor
XX
PS Claim 9; Fig 10; 180pp; English.
XX
CC The invention provides 22438, 23553, 25278 or 26212 human sulfatase
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides
CC are useful for treating disorders involving the brain such as
CC cerebrovascular diseases, infections such as acute meningitis,
CC demyelinating diseases including multiple sclerosis, degenerative
CC diseases affecting the cerebral cortex including Alzheimer's disease
CC and Pick disease, spinocerebellar degenerations including spinocerebellar
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,
CC degenerative diseases affecting motor neurons including amyotrophic
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,
CC toxic and acquired metabolic diseases, including vitamin deficiencies,
CC and neurocutaneous syndromes (phakamatoses) including neurofibromatosis.
CC The present sequence represents a human 25278 sulfatase polypeptide.
XX
XX Sequence 569 AA;

Query Match 3.5%; Score 19; DB 22; Length 569;
Best Local Similarity 100.0%; Pred. No. 2.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 GYSTHMVGKWHLGfYRKEC 186
Db 139 gysthmvgkwhlgfyrkcc 157

RESULT 10
AAB51184
ID AAB51184 standard; Protein; 533 AA.
XX
AC AAB51184;
XX
DT 21-MAR-2001 (first entry)
XX
DE Human sulfatase protein B SEQ ID NO:13.
XX
KW Human; iduronate 2-sulfatase; IDS; glycosylated; gene therapy;
KW mucopolysaccharidosis inhibitor; IDS deficiency disorder;
KW Hunter syndrome; mucopolysaccharidosis type II.
XX
OS Homo sapiens.
XX
PN US6153188-A.
XX
PD 28-NOV-2000.
XX
PF 12-FEB-1999; 99US-0249003.
XX
PR 17-DEC-1992; 92US-0991973.
PR 28-NOV-1994; 94US-0345212.
PR 12-NOV-1991; 91US-0790362.
XX
PA (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI Wilson PJ, Anson DS, Occhiodoro T, Bielicki J, Clements PR;
PI Hopwood JJ, Morris CP;
XX
DR WPI; 2001-060076/07.
XX
PT New highly glycosylated recombinant human iduronate 2-sulfatase (IDS)
PT useful for diagnosing or treating subjects suspected of having or
PT suffering from IDS deficiency disorders, e.g. Hunter syndrome
PT (mucopolysaccharidosis-II) -
XX
PS Example 1; Fig 3; 53pp; English.
XX
CC The present invention describes a recombinant human iduronate 2-sulfatase

CC (IDS). The recombinant IDS is more highly glycosylated than the naturally
 CC occurring enzyme isolated from human tissue. The recombinant human IDS
 CC can be produced in Chinese Hamster Ovary (CHO) cells or in a human cell.
 CC The recombinant IDS comprises a fusion protein. It is a
 CC mucopolysaccharidosis inhibitor and can be used in gene therapy.
 CC The recombinant IDS is useful in treating and diagnosing subjects
 CC suffering from or suspected of having IDS deficiency disorders, e.g.
 CC Hunter syndrome (mucopolysaccharidosis type II). The present sequence
 CC represents a human sulfatase which is used in comparison with the
 CC IDS sequence in an example from the present invention.

XX Sequence 533 AA;

Query Match 1.8%; Score 10; DB 22; Length 533;
 Best Local Similarity 100.0%; Pred. No. 0.37;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 THMVGKWHLG 180
 |||||
 Db 140 thmvgkwhlg 149

RESULT 11

ABB66123
 ID ABB66123 standard; Protein; 217 AA.

XX AC ABB66123;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 25161.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL10226.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX PS Disclosure; SEQ ID NO 25161; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 217 AA;

Query Match 1.5%; Score 8; DB 22; Length 217;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 ERVDSNR 511

|||||
 Db 35 ervdlsnr 42

RESULT 12

AAU36321
 ID AAU36321 standard; Protein; 330 AA.

XX AC AAU36321;

XX DT 14-FEB-2002 (first entry)

XX DE Pseudomonas aeruginosa cellular proliferation protein #311.

XX KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.

XX OS Pseudomonas aeruginosa.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX PR 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-20727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX DR WPI; 2001-611495/70.

XX DR N-PSDB; AAS54180.

XX PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -

XX PS Example 3; Seq ID No 11914; 511pp; English.

XX CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.

XX CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 330 AA;

Query Match 1.5%; Score 8; DB 22; Length 330;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 EEEGALLA 60
 |||||
 Db 171 eeegalla 178

RESULT 13

ABB63962
 ID ABB63962 standard; Protein; 542 AA.

XX AC ABB63962;

XX DT 26-MAR-2002 (first entry)

XX XX Drosophila melanogaster polypeptide SEQ ID NO 18678.

XX XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX XX 23-MAR-2001; 2001WO-US09231.

XX XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL08065.

XX XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX XX Disclosure; SEQ ID NO 18678; 2lpp + Sequence Listing; English.

XX XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 542 AA;

Query Match 1.5%; Score 8; DB 22; Length 542;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 PICTPSRS 127

|||||

Db 71 pictpsrs 78

RESULT 15

ABB63246

ID ABB63246 standard; Protein; 579 AA.

XX AC ABB63246;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 16530.

RESULT 14

AAB51186

ID AAB51186 standard; Protein; 551 AA.

XX AC AAB51186;

XX DT 21-MAR-2001 (first entry)

XX XX Sea urchin sulfatase protein U SEQ ID NO:15.

XX XX Human; iduronate 2-sulfatase; IDS; glycosylated; gene therapy;
 KW mucopolysaccharidosis inhibitor; IDS deficiency disorder;
 KW Hunter syndrome; mucopolysaccharidosis type II.

XX OS Arbacia punctulata.

XX PN US6153188-A.

XX PD 28-NOV-2000.

XX PF 12-FEB-1999; 99US-0249003.

XX PR 17-DEC-1992; 92US-0991973.

PR 28-NOV-1994; 94US-0345212.

PR 12-NOV-1991; 91US-0790362.

XX XX (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.

XX XX Wilison PJ, Anson DS, Occhiodoro T, Bielicki J, Clements PR;

PI Hopwood JJ, Morris CP;

XX XX WPI; 2001-060076/07.

XX XX New highly glycosylated recombinant human iduronate 2-sulfatase (IDS)
 PT useful for diagnosing or treating subjects suspected of having or
 PT suffering from IDS deficiency disorders, e.g. Hunter syndrome
 PT (mucopolysaccharidosis-II) -

XX XX Example 1; Fig 3; 53pp; English.

XX XX The present invention describes a recombinant human iduronate 2-sulfatase
 CC (IDS). The recombinant IDS is more highly glycosylated than the naturally
 CC occurring enzyme isolated from human tissue. The recombinant human IDS
 CC can be produced in Chinese Hamster Ovary (CHO) cells or in a human cell.
 CC The recombinant IDS comprises a fusion protein. It is a
 CC mucopolysaccharidosis inhibitor and can be used in gene therapy.
 CC The recombinant IDS is useful in treating and diagnosing subjects
 CC suffering from or suspected of having IDS deficiency disorders, e.g.
 CC Hunter syndrome (mucopolysaccharidosis type II). The present sequence
 CC represents a sea urchin sulfatase which is used in comparison with the
 CC IDS sequence in an example from the present invention.

XX XX Sequence 551 AA;

Query Match 1.5%; Score 8; DB 22; Length 551;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180

|||||

Db 153 mvgkwhlg 160

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
PF 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
PR (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX N-PSDB; ABL07349.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 16530; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 579 AA;

Query Match 1.5%; Score 8; DB 22; Length 579;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 174 VGKWHLGF 181
Db 127 vgkwhlglf 134
|||||||

Search completed: July 23, 2002, 19:08:41
Job time: 347 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:52:14 ; Search time 356.68 Seconds
(without alignments)
10907.615 Million cell updates/sec

Title: US-09-495-823-8
Perfect score: 2266
Sequence: 1 caccgctccgccacgcgtc.....tgccacctgtgccgaattc 2266

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1877	82.8	2253	AAH46863	Human 23553 sulfat
2	1549	68.4	1800	AAH46867	Human 23553 sulfat
3	1526	67.3	4639	AA545962	Human DNA encoding
4	1526	67.3	4640	AACT8483	Human PRO708 (UNQ3
5	1526	67.3	4650	AA233987	Human PRO708 nucle
6	890	39.3	1375	AA522726	Human cDNA encodin
7	590	26.0	1340	AA522490	Human cDNA encodin
8	537	23.7	562	AAH99356	Human protein enco
9	479	21.1	729	AAA02344	Human colon cancer

10	434	19.2	629	22	AA541205	cDNA encoding nove
11	55	2.4	294	21	AAA00278	Human colon cancer
12	44	1.9	1082	21	AAA02370	Human colon cancer
13	32	1.4	53	20	AAZ33990	Human PRO708 hybri
14	32	1.4	53	21	AAZ78675	Human immune/haema
15	26	1.1	774	22	AAK63748	Human polynucleoti
16	26	1.1	2029	22	AAK52385	Human cancer assoc
17	25	1.1	2613	21	AAK77614	cDNA encoding nove
18	25	1.1	2648	22	AAS34914	Human PRO708 PCR f
19	24	1.1	24	20	AAZ33988	Human PRO708 PCR f
20	24	1.1	24	20	AAZ33989	Human PRO708 PCR f
21	24	1.1	24	21	AAZ78673	Human PRO708 rever
22	24	1.1	24	21	AAZ78674	Human colon cancer
23	24	1.1	357	22	AAH36471	Human immune/haema
24	24	1.1	405	22	AAK57023	Human polynucleoti
25	24	1.1	457	22	AAI83302	Human secreted pro
26	24	1.1	1215	20	AAK00620	Apoptosis related
27	24	1.1	1316	21	AAK95796	Human immunoglobul
28	24	1.1	1325	22	AAZ28772	Human secreted pro
29	24	1.1	1403	22	AAZ05310	Human cDNA SEQ ID
30	24	1.1	1450	22	ABA06520	Human cDNA SEQ ID
31	24	1.1	1451	22	ABA06709	Human immunoglobul
32	24	1.1	1451	22	AAZ28838	Human immune/haema
33	24	1.1	1451	22	AAK61788	DNA encoding human
34	24	1.1	1578	22	AAZ33072	Human ADAM protein
35	24	1.1	2167	22	AAI81619	P. chrysogenum ABC
36	24	1.1	3133	22	AAF83398	Human genomic DNA
37	24	1.1	23041	22	AAZ26779	Human genomic DNA
38	24	1.1	23135	22	AAZ26780	Human immune/haema
39	24	1.1	23135	22	AAK81080	Human immune/haema
40	23	1.0	321	22	AAK56582	Human CDR62 homolo
41	23	1.0	359	22	ABA08227	Human polynucleoti
42	23	1.0	360	22	AAI90981	Human polynucleoti
43	23	1.0	378	22	AAI92024	Human polynucleoti
44	23	1.0	395	22	AAI32263	Human polynucleoti
45	23	1.0	398	22	AAI85187	Human polynucleoti

ALIGNMENTS

RESULT 1

AAH46863

ID AAH46863 standard; cDNA; 2253 BP.

AC AAH46863;

XX

XX 25-SEP-2001 (first entry)

DT

XX Human 23553 sulfatase polypeptide encoding cDNA.

DE

XX Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;

KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.

KW

XX Homo sapiens.

OS

XX Key

XX Location/Qualifiers

PH CDS

324..2123

FT

FT /**tag= a

FT

FT /product= "23553 sulfatase"

XX

XX WO200155411-A2.

XX

XX 02-AUG-2001.

XX

XX 31-JAN-2001; 2001WO-US03266.

XX

XX 31-JAN-2000; 2000US-0495823.

XX

XX (MILL-) MILLENNIUM PHARM INC.

XX

XX Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;

XX

Db 661 gactgtatgaagacaaatgctgctgggactatgacaatggcatatactccacag 720
QY 1044 atgtacactcagagtagcacgaaatcttagcttccataaaccacacaaagcctatatt 1103
Db 721 atgtacactcagagtagcacgaaatcttagcttccataaaccacacaaagcctatatt 780
QY 1104 ttatatattgctatcaagctgttcattccaccactgcaagctcctcggcaggtatttcgaa 1163
Db 781 ttatatattgctatcaagctgttcattccaccactgcaagctcctcggcaggtatttcgaa 840
QY 1164 cactaccgattccattatcaacataaacaggaggagatatgctgcccattgcttccctgctta 1223
Db 841 cactaccgattccattatcaacataaacaggaggagatatgctgcccattgcttccctgctta 900
QY 1224 gatgaagcaatacaacaagtgacattgctctaaagacttatgggtttctataacaacagc 1283
Db 901 gatgaagcaatacaacaagtgacattgctctaaagacttatgggtttctataacaacagc 960
QY 1284 attatcatttactcttcagataatggtgcccagctacgagcagggagtaactggcct 1343
Db 961 attatcatttactcttcagataatggtgcccagctacgagcagggagtaactggcct 1020
QY 1344 ctacaggttagcaaggagacaatttgggaaggaggatccggcctgtaggctttgtgcat 1403
Db 1021 ctacaggttagcaaggagacaatttgggaaggaggatccggcctgtaggctttgtgcat 1080
QY 1404 agcccactctgaagaaacaggaaacagtggttaaggaacctgtgcacatacactgactgg 1463
Db 1081 agcccactctgaagaaacaggaaacagtggttaaggaacctgtgcacatacactgactgg 1140
QY 1464 taccctactctatttccactggtggaagacagattgatgaggaacctcaactagatggc 1523
Db 1141 taccctactctatttccactggtggaagacagattgatgaggaacctcaactagatggc 1200
QY 1524 tatgatattgggaagacaataagtgagggtcttcgctcaccctcagtagatatttgcatt 1583
Db 1201 tatgatattgggaagacaataagtgagggtcttcgctcaccctcagtagatatttgcatt 1260
QY 1584 aaattgaccccatatatacaccagcaaaaaatgctcctcggcagcaggtatggatc 1643
Db 1261 aaattgaccccatatatacaccagcaaaaaatgctcctcggcagcaggtatggatc 1320
QY 1644 tggaaactgcgaatcagtcagccatcagagtgagcagcagtgaaattgcttacaggaat 1703
Db 1321 tggaaactgcgaatcagtcagccatcagagtgagcagcagtgaaattgcttacaggaat 1380
QY 1704 cctggtacagcagactgggtcccccctcagctcttcagcaaacctgggaccgacccggtgg 1763
Db 1381 cctggtacagcagactgggtcccccctcagctcttcagcaaacctgggaccgacccggtgg 1440
QY 1764 caaatgaacggtacactcgtcaactggcaaaagtatgctgtttcaacatacagcc 1823
Db 1441 caaatgaacggtacactcgtcaactggcaaaagtatgctgtttcaacatacagcc 1500
QY 1824 gaccatattgaaggtggagccttatcaccaggtatccaggaatcgtgaagagctccta 1883
Db 1501 gaccatattgaaggtggagccttatcaccaggtatccaggaatcgtgaagagctccta 1560
QY 1884 cggaggtctcaggttcaacaaaaactgcagtgccgggtcaggtatccccccaaagacccc 1943
Db 1561 cggaggtctcaggttcaacaaaaactgcagtgccgggtcaggtatccccccaaagacccc 1620
QY 1944 aqaagtaacctgagctcaatggagggatct 1974
Db 1621 aqaagtaacctgagctcaatggagggatct 1651

RESULT 3
AAS45962
ID: AAS45962 standard; cDNA; 4639 BP.
XX
AC AAS45962;

XX
DT 18-DEC-2001 (first entry)
XX
DE Human DNA encoding PRO polypeptide sequence #38.
XX
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168848-A2.
XX
PD 20-SEP-2001.
XX
PF 28-FEB-2001; 2001WO-US06520.
XX
PR 01-MAR-2000; 2000WO-US05601.
PR 02-MAR-2000; 2000WO-US05841.
PR 03-MAR-2000; 2000US-187202P.
PR 06-MAR-2000; 2000US-186968P.
PR 14-MAR-2000; 2000US-189320P.
PR 14-MAR-2000; 2000US-189320P.
PR 15-MAR-2000; 2000WO-US06884.
PR 21-MAR-2000; 2000US-190828P.
PR 21-MAR-2000; 2000US-191007P.
PR 21-MAR-2000; 2000US-191048P.
PR 21-MAR-2000; 2000US-191314P.
PR 28-MAR-2000; 2000US-192655P.
PR 29-MAR-2000; 2000US-193032P.
PR 30-MAR-2000; 2000US-193053P.
PR 30-MAR-2000; 2000WO-US08439.
PR 04-APR-2000; 2000US-194449P.
PR 04-APR-2000; 2000US-194647P.
PR 11-APR-2000; 2000US-195975P.
PR 11-APR-2000; 2000US-196000P.
PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199550P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
(GETH) GENENTECH INC.
XX
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI: 2001-602746/68.
DR P-PSDB; AAU29061.
XX
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
XX Claim 2; Fig 75; 774pp; English.
XX

QY 1162 aactaccgattccattatcaacataaacagagagagagatgctgcttctgct 1221
DB 1729 aactaccgattccattatcaacataaacagagagagatgctgcttctgct 1788
QY 1222 tagatgaagcaatcaacaaacgtgacattgctctaaagacttatggtttctataacaaca 1281
DB 1789 tagatgaagcaatcaacaaacgtgacattgctctaaagacttatggtttctataacaaca 1848
QY 1282 gcattatcattactcttcagataatggtgccagctacgagcagagggagtaactggc 1341
DB 1849 gcattatcattactcttcagataatggtgccagctacgagcagagggagtaactggc 1908
QY 1342 ctctcagaggtagcaaaagaaacatatgttggaagggaggtccggctgaggtcttggc 1401
DB 1909 ctctcagaggtagcaaaagaaacatatgttggaagggaggtccggctgaggtcttggc 1968
QY 1402 atagcccactcttgaaaaaaggaacagtgtgttaagaaacctgtgcacatcaactgact 1461
DB 1969 atagcccactcttgaaaaaaggaacagtgtgttaagaaacctgtgcacatcaactgact 2028
QY 1462 ggtaccccaactctattcactggtcgaagacagagattgatgagacattcaactagatg 1521
DB 2029 ggtaccccaactctattcactggtcgaagacagagattgatgagacattcaactagatg 2088
QY 1522 gctatgatctggagaccataagtgggtcttcgctcaccgcagagtagatatgttgc 1581
DB 2089 gctatgatctggagaccataagtgggtcttcgctcaccgcagagtagatatgttgc 2148
QY 1582 ataacattgacccc-tatacaccaggcaaaaaatggctcctgggcagcaggtatggga 1641
DB 2149 ataacattgacccc-tatacaccaggcaaaaaatggctcctgggcagcaggtatggga 2207
QY 1642 tctggaacactgcaatccagtcagccatcagagtcagactggaattgcttacagaaa 1701
DB 2208 tctggaacactgcaatccagtcagccatcagagtcagactggaattgcttacagaaa 2267
QY 1702 atcttggtacagcagactgggttccccctcagttcttccagaaacctgggaccgaccggt 1761
DB 2268 atcttggtacagcagactgggttccccctcagttcttccagaaacctgggaccgaccggt 2327
QY 1762 ggcacaaatgaacggatccctgtaacctgcaacgaaagtgtatggttttcaacatcaacg 1821
DB 2328 ggcacaaatgaacggatccctgtaacctgcaacgaaagtgtatggttttcaacatcaacg 2387
QY 1822 ccgacccatagagaggtgacctatcaacaggtatccaggaatcgtgaagaagctcc 1881
DB 2388 ccgacccatagagaggtgacctatcaacaggtatccaggaatcgtgaagaagctcc 2447
QY 1882 tacggaggctctcacagttcaacaaaactgcagtgccggtcagggtatccccccaaagacc 1941
DB 2448 tacggaggctctcacagttcaacaaaactgcagtgccggtcagggtatccccccaaagacc 2507
QY 1942 ccagaagttaacctagggtcaatgaggggtct 1974
DB 2508 ccagaagttaacctagggtcaatgaggggtct 2540

RESULT 6

ID AAS22726 standard; cDNA; 1375 BP.

XX

AC AAS22726;

XX

DT 24-OCT-2001 (first entry)

XX

DE Human cDNA encoding a novel human protein #292.

XX

KW Human; novel protein; ss; Antianemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; anti allergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;

KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.

OS Homo sapiens.

XX WO200155437-A2.

PN 02-AUG-2001.

XX

PF 25-JAN-2001; 2001WO-US02623.

XX

PR 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451939/48.

DR P-PSDB; AAU14421.

XX

PT Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage -

PS Claim 1; Page 699; 894pp; English.

XX

The invention relates to polynucleotides encoding novel human
proteins or their active domains. The polypeptides, polynucleotides and
antibodies raised against the polypeptides are used in a method of
treatment of a mammal and prevention of disorders caused by the aberrant
protein expression or activity. The polypeptides can be used as
molecular weight markers, food supplements, and in antibody production.
The polypeptides are used to identify compounds which bind to the
polypeptides. Polynucleotides of the invention are used as probes and
primers for sequencing, for chromosome or gene mapping, in the
production of recombinant proteins, and in generating anti-sense DNA or
RNA and in gene therapy. Polypeptides of the invention can be used to
target drugs to a tumour, in assays to determine biological activity, to
raise antibodies/elicit an immune response, to determine quantitative
protein levels, as tissue markers, and to isolate receptors or ligands.
Polypeptides of the invention may also be useful in treating platelet
disorders, stem cell disorders, regenerating bone, cartilage, tendon,
ligament and/or nerve tissue, wound healing, treating burns, promoting
the proliferation, differentiation and survival of stem cells, as a
contraceptive, treating osteoporosis and osteoarthritis, anaemia,
Alzheimer's, parkinson's and Huntington's diseases, anlytrophic lateral
sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
fungal infection or from autoimmunity, cancer, allergy, asthma,
graft-versus-host disease, eczema, haemophilia, thrombosis,
anti-inflammatory diseases, nervous system disorders, and infection.
The present sequence encodes a protein of the invention.

XX Sequence 1375 BP; 399 A; 336 C; 360 G; 280 T; 0 other;

Query Match 39.3%; Score 890; DB 22; Length 1375;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 890; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 caggcctgtgtctctctggaagatgtagcaatggggcgctggcagattctggatc 428
DB 484 caggcctgtgtctctctggaagatgtagcaatggggcgctggcagattctggatc 543

QY 429 ctctgcctctcacttatggttacctgtctctggggccaggccttagaagaggaagaa 488

DB 544 ctctgcctctcacttatggttacctgtctctggggccaggccttagaagaggaagaa 603

QY 489 ggggccttactagctcaagctggagaaactagaccagcacacttcccactcccc 548

DB 604 ggggccttactagctcaagctggagaaactagaccagcacacttcccactcccc 663

QY 549 cccatctcatttcatctcctcagcagatcatcagggaatttagagatgtgggttaccacgga 608

DB 664 cccatctcatttcatctcctcagcagatcatcagggaatttagagatgtgggttaccacgga 723

QY 609 tctgagattaaacacccactactcttgacaagctcgtcgccgaaggaggttaaaactggagac 668
DB 724 tctgagattaaacacccactactcttgacaagctcgtcgccgaaggaggttaaaactggagac 783
QY 669 tactatgtccagcctatttgacacaccatccaggagtcagttatttactggaaagtacag 728
DB 784 tactatgtccagcctatttgacacaccatccaggagtcagttatttactggaaagtacag 843
QY 729 atacacacggagcttcaacattctatcataagacacctacccaacaaactgtttactcttg 788
DB 844 atacacacggagcttcaacattctatcataagacacctacccaacaaactgtttactcttg 903
QY 789 gacaatgcacccctaccctcagaaaactgaagggttgatattcaacgcataatggtcgga 848
DB 904 gacaatgcacccctaccctcagaaaactgaagggttgatattcaacgcataatggtcgga 963
QY 849 aaatggcacttgggtttttacagaaaaaatgcatgccacacagaaaggaggtttgtatacc 908
DB 964 aaatggcacttgggtttttacagaaaaaatgcatgccacacagaaaggaggtttgtatacc 1023
QY 909 tttttgttcccttttgggaagtgggattactactatcacactacaaatgtgacagtctc 968
DB 1024 tttttgttcccttttgggaagtgggattactactatcacactacaaatgtgacagtctc 1083
QY 969 gggatgtgggtatgactgttatgaaaacgacaaatgctgctgggactatgacaaatggc 1028
DB 1084 gggatgtgggtatgactgttatgaaaacgacaaatgctgctgggactatgacaaatggc 1143
QY 1029 atatactccacacagatgtacactcagagagtacagcaaatettatgcttccataacccc 1088
DB 1144 atatactccacacagatgtacactcagagagtacagcaaatettatgcttccataacccc 1203
QY 1089 acaagcctatttttatatttcctcatcaagctgttcttaccactgcaagctcct 1148
DB 1204 acaagcctatttttatatttcctcatcaagctgttcttaccactgcaagctcct 1263
QY 1149 ggcaggtatttcgaacactaccgattccattatcaacataaacagagagatgctgccc 1208
DB 1264 ggcaggtatttcgaacactaccgattccattatcaacataaacagagagatgctgccc 1323
QY 1209 atgcttctcgtcttagatgaagcaatacaacagtcgacattggctctaaa 1258
DB 1324 atgcttctcgtcttagatgaagcaatacaacagtcgacattggctctaaa 1373

RESULT 7
AAS22490
ID AAS22490 standard; cDNA; 1340 BP.
XX
AC AAS22490;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human cDNA encoding a novel human protein #56.
XX
KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-451939/48.
DR P-PSDB; AAU14185.
XX
XX Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage -
PS Claim 1; Page 256-257; 894pp; English.
XX
XX The invention relates to polynucleotides encoding novel human
XX proteins or their active domains. The polypeptides, polynucleotides and
XX antibodies raised against the polypeptides are used in a method of
XX treatment of a mammal and prevention of disorders caused by the aberrant
XX protein expression or activity. The polypeptides can be used as
XX molecular weight markers, food supplements, and in antibody production.
XX The polypeptides are used to identify compounds which bind to the
XX polypeptides. Polynucleotides of the invention are used as probes and
XX primers, for sequencing, for chromosome or gene mapping, in the
XX production of recombinant proteins, and in generating anti-sense DNA or
XX RNA and in gene therapy. Polypeptides of the invention can be used to
XX target drugs to a tumour, in assays to determine biological activity, to
XX raise antibodies/ elicit an immune response, to determine quantitative
XX protein levels, as tissue markers, and to isolate receptors or ligands.
XX Polypeptides of the invention may also be useful in treating platelet
XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX ligament and/or nerve tissue, wound healing, treating burns, promoting
XX the proliferation, differentiation and survival of stem cells, as a
XX contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX fungal infection or from autoimmunity, cancer, allergy, asthma,
XX graft-versus-host disease, eczema, haemophilia, thrombosis,
XX anti-inflammatory diseases, nervous system disorders, and infection.
XX The present sequence encodes a protein of the invention.
SQ Sequence 1340 BP; 374 A; 338 C; 369 G; 258 T; 1 Other;

Query Match 26.0%; Score 590; DB 22; Length 1340;
Best Local Similarity 100.0%; Pred. No. 2.6e-272;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 369 caggcctgtgtcctcggaagatgtagcaatggggcgctggcaggattctggatc 428
DB 484 caggcctgtgtcctcggaagatgtagcaatggggcgctggcaggattctggatc 543
QY 429 ctctgcctcctcaactatgttacctgtcctggggccaggccttagagagaggagaa 488
DB 544 ctctgcctcctcaactatgttacctgtcctggggccaggccttagagagaggagaa 603
QY 489 ggggccttactagctcaagctggagaaaactagagccagcacaaactccacctccag 548
DB 604 ggggccttactagctcaagctggagaaaactagagccagcacaaactccacctccag 663
QY 549 ccccatctcatttccatcctcagcgatgatcaggatttagagatgtgggttacacagga 608
DB 664 ccccatctcatttccatcctcagcgatgatcaggatttagagatgtgggttacacagga 723
QY 609 tctgagattaaaaaaccttactcttgacaagctcgtcgccgaaggaggttaaaactggagac 668
DB 724 tctgagattaaaaaaccttactcttgacaagctcgtcgccgaaggaggttaaaactggagac 783
QY 669 tactatgtccagcctatttgacacaccatccaggagtcagttatttactggaaagtacag 728
DB 784 tactatgtccagcctatttgacacaccatccaggagtcagttatttactggaaagtacag 843
QY 729 atacacacggagcttcaacattctatcataagacacctacccaacaaactgtttactcttg 788
DB 844 atacacacggagcttcaacattctatcataagacacctacccaacaaactgtttactcttg 903

Qy 789 gacaatgccaccctacatcagaaactgaagaggttgatattcaacgcataatggtcgga 848
|||||
Db 904 gacaatgccaccctacatcagaaactgaagaggttgatattcaacgcataatggtcgga 963

Qy 849 aaatggcaactgggtttttacagaaaagaatgcagtcgccaccagagaagagattgatacc 908
|||||
Db 964 aaatggcaactgggtttttacagaaaagaatgcagtcgccaccagagaagagattgatacc 1023

Qy 909 tttttggtcccttttgggaagtgggattactatcacacactacaagt 958
|||||
Db 1024 tttttggtcccttttgggaagtgggattactatcacacactacaagt 1073

RESULT 8
AAH99356
ID AAH99356 standard; cDNA; 562 BP.
XX
AC AAH99356;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human protein encoding cDNA sequence SEQ ID NO:191.
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder; ss.

OS Homo sapiens.
XX
XX WO200153455-A2.
XX
XX 26-JUL-2001.
XX
XX 22-DEC-2000; 2000WO-US35017.
XX
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-457603/49.
DR P-PSDB; AAM25415.
XX
XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX
XX Claim 1; Page 383; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAH25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides

CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX

SEQ Sequence 562 BP; 170 A; 132 C; 105 G; 155 T; 0 other;

Query Match 23.7%; Score 537; DB 22; Length 562;
Best Local Similarity 100.0%; Pred. NO. 6.4e-247;
Matches 537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 722 gtatcagatcacaccggacttcaacattctatcaagacactcccaacccaactgttt 781
|||||
Db 24 gtatcagatcacaccggacttcaacattctatcaagacactcccaacccaactgttt 83

Qy 782 acctgtgacaatgccaccctacctcagaaactgaagaggttgatattcaacgcata 841
|||||
Db 84 acctgtgacaatgccaccctacctcagaaactgaagaggttgatattcaacgcata 143

Qy 842 ggtcggaaaaatggcacttgggtttttacagaaaagaatgcagtcgccaccagaagagatt 901
|||||
Db 144 ggtcggaaaaatggcacttgggtttttacagaaaagaatgcagtcgccaccagaagagatt 203

Qy 902 tgatacctttttgttcccttttgggaagtgggattactatcacactacaaatgtga 961
|||||
Db 204 tgatacctttttgttcccttttgggaagtgggattactatcacactacaaatgtga 263

Qy 962 cagtcctgggagtgtggtgctagtgtatgaaaaacgacaatgctgctgggactatga 1021
|||||
Db 264 cagtcctgggagtgtggtgctagtgtatgaaaaacgacaatgctgctgggactatga 323

Qy 1022 caatggcatatactccacacagatgtacactcagagatcacagaaatcttagcttccca 1081
|||||
Db 324 caatggcatatactccacacagatgtacactcagagatcacagaaatcttagcttccca 383

Qy 1082 taaccccaaaagcctatatattttatatattgctctataaagctgttctaccactgca 1141
|||||
Db 384 taaccccaaaagcctatatattttatatattgctctataaagctgttctaccactgca 443

Qy 1142 agctcctggcagggtatttcgaacactaccgattccattatcaacataacagaggagata 1201
|||||
Db 444 agctcctggcagggtatttcgaacactaccgattccattatcaacataacagaggagata 503

Qy 1202 tgetgccaatgcttccctgcttagatgaagaacatacaacacgctgacattggctctaaa 1258
|||||
Db 504 tgetgccaatgcttccctgcttagatgaagaacatacaacacgctgacattggctctaaa 560

RESULT 9
AAH02344
ID AAH02344 standard; cDNA; 729 BP.
XX
AC AAH02344;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2335.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.

XX WO9958675-A2.
PN 18-NOV-1999.
PD 13-MAY-1999; 99WO-US10602.
PF 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.
DR Polynucleotide library used to determine cancerous states of mammalian
XX cells -
PT
PT
PS Claim 1; Page 924-925; 1097pp; English.
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX Sequence 729 BP; 232 A; 178 C; 163 G; 143 T; 13 other;
SQ

Query Match 21.1%; Score 479; DB 21; Length 729;
Best Local Similarity 99.8%; Pred. No. 3.9e-219;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1445 tgtgcacatcaactgactggtaccacactctcttcaactggtgacgacagattgatga 1504
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
47 tgtgcacatcaactgactggtaccacactctcttcaactggtgacgacagattgatga 106
QY 1505 ggacattcaactagatgctatgatctatctggagaccataagtgagggtcttcgtcacc 1564
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
107 ggacattcaactagatgctatgatctggtgagaccataagtgagggtcttcgtcacc 166
QY 1565 ccgagtagatatttgcataacatgaccccatatcacaccaggccaaaatggctcctg 1624
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
167 ccgagtagatatttgcataacatgaccccatatcacaccaggccaaaatggctcctg 226
QY 1625 ggcagcaggctatggatctggaacactgcaatccagtcagccatcagatgagcagctg 1684
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
227 ggcagcaggctatggatctggaacactgcaatccagtcagccatcagatgagcagctg 286
QY 1685 gaattgtcttacagaaaatcctggtctacagcagactgggtcccccctcagttttcagcaa 1744
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
287 gaattgtcttacagaaaatcctggtctacagcagactgggtcccccctcagttttcagcaa 346

QY 1745 cctgggaccgaaccggtggcacaatgaacgatcacctcgtcaactggcacaagtgtatg 1804
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
347 cctgggaccgaaccggtggcacaatgaacgatcacctcgtcaactggcacaagtgtatg 406
QY 1805 gcttttcaacatcacagccgaccccatatgagagggtggacctatctaacaggtatccag 1864
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
407 gcttttcaacatcacagccgaccccatatgagagggtggacctatctaacaggtatccag 466
QY 1865 aatcgtgaagaagctccttacggaggtctctcacagttcaacaaaactgcagtcgccggtcag 1924
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
467 aatcgtgaagaagctccttacggaggtctctcacagttcaacaaaactgcagtcgccggtcag 526
QY 1925 gtatcccccccaagaagcccaagaagtaaccctaggtcctcaatggagggtct 1974
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
527 gtatcccccccaagaagcccaagaagtaaccctaggtcctcaatggagggtct 576

RESULT 10
AAS41205
ID AAS41205 standard; cDNA; 629 BP.
XX
AC AAS41205;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #421.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0244618.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465566/50.
DR P-PSDB; AAU23335.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Claim 4; SEQ ID No 431; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU40785-AAU41684 represent cDNA sequences encoding for the novel human
CC enzyme polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 629 BP; 189 A; 151 C; 124 G; 162 T; 3 other;

Query Match 19.2%; Score 434; DB 22; Length 629;
Best Local Similarity 99.8%; Pred. No. 1.4e-197;
Matches 554; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 589 gagatgtgggttacacgagatcgagattaaaacacacactactcttgacaaagctcgctgcg 648
|||||
Db 14 gagatgtgggttacacgagatcgagattaaaacacactactcttgacaaagctcgctgcg 73
|||||
QY 649 aaggagttaactggagaaactactatgtctccagcctatttgcacaccatccaggagtcagt 708
|||||
Db 74 aaggagttaactggagaaactactatgtctccagcctatttgcacaccatccaggagtcagt 133
|||||
QY 709 ttattactggaagtatcagatatacacacgcggaacttcaacattctctatcataagacctaacc 768
|||||
Db 134 ttattactggaagtatcagatatacacacgcggaacttcaacattctctatcataagacctaacc 193
|||||
QY 769 aaccacaaactgtttactctctgacaaatgccaccctacccagaaaactgaagaggttgat 828
|||||
Db 194 aaccacaaactgtttactctctgacaaatgccaccctacccagaaaactgaagaggttgat 253
|||||
QY 829 attcaacgcataatgctcggaataatggcacttgggtttttacagaaaagaatgcacgccca 888
|||||
Db 254 attcaacgcataatgctcggaataatggcacttgggtttttacagaaaagaatgcacgccca 313
|||||
QY 889 ccagaagaggatttgatacctttttt-ggtcccttttgggaagtgggattactatataca 947
|||||
Db 314 ccagaagaggatttgataccttttttgggtcccttttgggaagtgggattactatataca 373
|||||
QY 948 cactacaaatgtgacagtcctcggtggtggtgtggtctatgactgtatgaaaacgacaatgct 1007
|||||
Db 374 cactacaaatgtgacagtcctcggtggtggtgtggtctatgactgtatgaaaacgacaatgct 433
|||||
QY 1008 gcttgggactatgacaatggcatactactccacacagatgtacactcagagagtagacgaa 1067
|||||
Db 434 gcttgggactatgacaatggcatactactccacacagatgtacactcagagagtagacgaa 493
|||||
QY 1068 attctagttccctataaccccaagaagcctatatattttatatatttgcctatacaagctgtt 1127
|||||
Db 494 attctagttccctataaccccaagaagcctatatattttatatatttgcctatacaagctgtt 553
|||||
QY 1128 cattcaccactgcaa 1142
|||||
Db 554 cattcaccactgcaa 568
|||||

RESULT 11
AAA00278
ID AAA00278 standard; cDNA: 294 BP.
XX
AC AAA00278;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:269.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
FA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.
XX Polynucleotide library used to determine cancerous states of mammalian
PI cells -
XX
PS Claim 1; Page 244-245; 1097pp; English.
XX
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC libraries constructed from human colon cancer cell lines. The present
CC invention also describes a method of detecting differentially expressed
CC genes correlated with a cancerous state of a mammalian cell, comprising
CC detecting at least one differentially expressed gene product in a test
CC sample derived from a cell suspected of being cancerous, where detection
CC of the differentially expressed gene product is correlated with a
CC cancerous state of the cell from which the test sample was derived.
CC The polynucleotides sequences can be used in a method for detecting
CC differentially expressed genes correlated with a cancerous state of a
CC mammalian cell. The polynucleotides can also be used as probes for
CC detecting and mapping related genes. They can be used in diagnosis and
CC prognosis of diseases and disorders (e.g. identification of
CC pre-metastatic or metastatic cancerous states, stages of cancer, or
CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 294 BP; 79 A; 46 C; 51 G; 102 T; 16 other;

Query Match 2.4%; Score 55; DB 21; Length 294;
Best Local Similarity 100.0%; Pred. No. 5.4e-16;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1481 actggtgaagacagattgatgagacattcaactagatgcttatgatctgg 1535
Db 65 actggtgaagacagattgatgagacattcaactagatgcttatgatctgg 119
|||||

RESULT 12
AAA02370
ID AAA02370 standard; cDNA: 1082 BP.
XX
AC AAA02370;
XX
DT 19-MAY-2000 (first entry)
XX
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2361.
XX
KW Human; colon cancer; tumour; diagnosis; gene expression product;
KW probe; detection; cancerous state; metastasis; identification;
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9958675-A2.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US10602.
XX
PR 14-MAY-1998; 98US-0085426.
PR 15-MAY-1998; 98US-0085537.
PR 15-MAY-1998; 98US-0085696.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.


```
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX (GETH ) GENENTECH INC.
XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX WPI; 1999-551358/46.
XX
XX New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
XX Example 19; Page 197; 530pp; English.
XX
XX The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ34338, and AA41685 to AA41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
XX invention.
XX
SQ Sequence 53 BP; 14 A; 13 C; 14 G; 11 T; 1 other;

Query Match 1.4%; Score 32; DB 20; Length 53;
Best Local Similarity 100.0%; Pred. No. 5.6e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 gccaccctacctcagaaactgaaggaggttg 826
Db 1 gccaccctacctcagaaactgaaggaggttg 32

RESULT 14
AAC78675
ID AAC78675 standard; DNA; 53 BP.
XX
AC AAC78675;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PRO708 hybridisation probe SEQ ID NO:117.
XX
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW expressed sequence tag; detection; cancer; PCR primer; probe; ss.
XX
OS Homo sapiens.
XX
FN WO200053756-A2.
XX
PD 14-SEP-2000.
XX
PF 18-FEB-2000; 2000WO-US04341.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 14-MAY-1999; 99US-0134287.
```

```
PR 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0145698.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fillvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
XX WPI; 2000-611443/58.
XX
XX Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX
XX Example 19; Page 250; 536pp; English.
XX
XX AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
SQ Sequence 53 BP; 14 A; 13 C; 14 G; 11 T; 1 other;

Query Match 1.4%; Score 32; DB 21; Length 53;
Best Local Similarity 100.0%; Pred. No. 5.6e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 gccaccctacctcagaaactgaaggaggttg 826
Db 1 gccaccctacctcagaaactgaaggaggttg 32

RESULT 15
AAK63748
ID AAK63748 standard; cDNA; 774 BP.
XX
AC AAK63748;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding CDNA SEQ ID NO:8808.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
FN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
```


XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0218290.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX P-PSDB; AAM90967.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:12:04 ; Search time 4094.62 Seconds
(without alignments)
11580.936 Million cell updates/sec

Title: US-09-495-823-8
Perfect score: 2266
Sequence: 1 cagcgctcgccacgcgtc.....tgcacctggtgcgaattc 2266

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- GenEmbl: *
- 1: gb_ba:*
 - 2: gb_htg:*
 - 3: gb_in:*
 - 4: gb_om:*
 - 5: gb_ov:*
 - 6: gb_pat:*
 - 7: gb_ph:*
 - 8: gb_pl:*
 - 9: gb_pr:*
 - 10: gb_ro:*
 - 11: gb_sts:*
 - 12: gb_sy:*
 - 13: gb_un:*
 - 14: gb_vi:*
 - 15: em_ba:*
 - 16: em_fun:*
 - 17: em_hum:*
 - 18: em_in:*
 - 19: em_mu:*
 - 20: em_om:*
 - 21: em_or:*
 - 22: em_ov:*
 - 23: em_pat:*
 - 24: em_ph:*
 - 25: em_pl:*
 - 26: em_ro:*
 - 27: em_sts:*
 - 28: em_un:*
 - 29: em_vi:*
 - 30: em_htg_hum:*
 - 31: em_htg_inv:*
 - 32: em_htg_other:*
 - 33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	1877	82.8	2253	6	AX206967
2	1549	68.4	1800	6	AX206973
c 3	1207	53.3	152861	9	AC104779 Homo sapi
c 4	1207	53.3	165147	2	AC021342 Homo sapi
c 5	630	27.8	157043	2	AC013692 Homo sapi
c 6	630	27.8	165147	2	AC021342 Homo sapi
c 7	630	27.8	175645	2	AC105414 Homo sapi
c 8	263	11.6	1871	9	AK027201 Homo sapi
c 9	99	4.4	176341	2	AC105383 Homo sapi
c 10	53	2.3	232951	2	AC091322 Mus muscu
c 11	29	1.3	122992	2	AC098604 Rattus no
c 12	26	1.1	927	9	BC009528 Homo sapi
c 13	26	1.1	2003	10	BC012659 Mus muscu
c 14	25	1.1	28735	2	AC108339 Rattus no
c 15	25	1.1	38866	3	CBRG36E21 Caenorhab
c 16	25	1.1	191744	2	AC044865 Homo sapi
c 17	25	1.1	211652	9	AP001838 Homo sapi
c 18	24	1.1	1603	10	RATARSB D49434 Rat ARSB mr
c 19	24	1.1	2491	10	BC021406 Mus muscu
c 20	24	1.1	3695	10	BC021427 Mus muscu
c 21	24	1.1	118722	2	AP000584 Homo sapi
c 22	24	1.1	136503	2	AC068939 Homo sapi
c 23	24	1.1	145718	2	AC091369 Rattus no
c 24	24	1.1	153400	2	AC106436 Rattus no
c 25	24	1.1	153704	9	AC008567 Homo sapi
c 26	24	1.1	153882	2	AC011256 Homo sapi
c 27	24	1.1	161591	2	AP001142 Homo sapi
c 28	24	1.1	170537	2	AC073548 Homo sapi
c 29	24	1.1	173435	2	AP002386 Homo sapi
c 30	24	1.1	174484	9	AP002815 Homo sapi
c 31	24	1.1	176852	2	AC097993 Rattus no
c 32	24	1.1	180672	9	AC011451 Homo sapi
c 33	24	1.1	191919	9	AC016770 Homo sapi
c 34	24	1.1	219366	9	AP001922 Homo sapi
c 35	23	1.0	482	6	AR128083 Sequence
c 36	23	1.0	1216	6	AX277154 Sequence
c 37	23	1.0	1540	9	AJ420512 Homo sapi
c 38	23	1.0	2954	10	BC021508 Mus muscu
c 39	23	1.0	3451	5	AF072455 Xenopus l
c 40	23	1.0	4096	9	BC015055 Homo sapi
c 41	23	1.0	4858	6	AX347398 Sequence
c 42	23	1.0	4858	6	AX349119 Sequence
c 43	23	1.0	4918	10	BC021490 Mus muscu
c 44	23	1.0	17514	2	AC012851 Drosophil
c 45	23	1.0	20280	9	AL671136 Human DNA

ALIGNMENTS

RESULT 1	AX206967	2253 bp	DNA	linear	PAT 30-AUG-2001
LOCUS	Sequence 8 from Patent WO0155411.				
DEFINITION	AX206967				
ACCESSION	AX206967.1	GI:15394724			
VERSION					
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2253)				
AUTHORS	Glucksmann, M.A., Williamson, M., Rudolph-Owen, L.A. and Tsai, F.Y.				
TITLE	Human sulfatases				
JOURNAL	Patent: WO 0155411-A 8 02-AUG-2001;				
FEATURES	Millennium Pharmaceuticals, Inc. (US)				
source	Location/Qualifiers				
	1..2253				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	324..2123				
CDS	/note="unnamed protein product"				
	/codon_start=1				

/protein_id="CAC60191.1"
/db_xref="GI:15394725"
/translation="MAPRGAGHPPPPSPPOACVCPGKMLAMGALAGRWILCLLTGYVL
SWGOALEEEEGALLAQAGEKLEPSTSTSOPLHILFILDADQGFROVYHGSEIKPT
LDKLAEGVKLENYVQIPTCPSRQFITYQIHTGLQHSIIRPQPNCLPDNATL
PQKLKEVGYSTHVMGWHLGFYKRCMPTRRGEDTFGSLGSDYTHYKCDSPGMC
GYDLYENDNAWDYDNGIY2QMYTORVOQILASHNPKPIFLYIAFYVHSLQAPG
RYFEHYRINIINRRYAAIMSLCLDEAINNVIALKTYGFYNNISIIYSDNGQOPTA
GGSNPLRGSKGYWEGGIRAVGFVHSPLLKNKGTVCVELVHITDWPPTLISLAEGOI
DEDIQLDGYDMEITSEGLSPRVIDLHNIDPIYTKAKNGSWAAGGIWNTAIOQSAIR
VQHWKLDTGNPGFSDWVPPOSFNGLGNRWHNERITLSTGKSVLFINITADPIYERVDL
SNRPYGVIKKLLRLSOFNKTAVPYRPKDPDRSNRLNGVMPGPMYKEETKKKKPSK
NQAEKKQSKKKKKQKQKAVSGTSCHSYTCG"
BASE COUNT 624 a 535 c 554 g 540 t
ORIGIN

Query Match 82.8%; Score 1877; DB 6; Length 2253;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2247; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
Qy 1 caecggtccgcacggtccgtgagagataataactttttttttttttttttttttttttt 60
Db 1 CACGCGTCCGCCACGCGTCCGTGGAGATATAACTTTTCTTTTCTTTTCTTCTTGGT 60
Qy 61 ggaagctgcttagggagggggagggagagagaaagtgaatgctgctggagaagagcg 120
Db 61 GGAAGCTGCTTAGGAGGGGGAGGAGGAGGAGAAAGTGAANTGCTGTGAGAGAGCG 120
Qy 121 agccctcctgtttccggagtcocattccatttaagccatcacttctggaagataaagt 180
Db 121 AGCCCTCCTTCTTCGGAGTCCCATCCATTAAAGCCATCACTTCTGGAAGATTAAGT 180
Qy 181 tgcggacatggtcacagctagagagagagagagatcttctgccaggtggagagcttca 240
Db 181 TGTGGACATGGTGACAGCTGAGAGGAGAGAGAGATTTCTTGCCAGGTGGAGAGTCTTCA 240
Qy 241 ccgtctgttgggtgcatgtgtgcgccgcacscggcgggcggtggttctccgcgtgg 300
Db 241 CCGTCTGTTGGTGATGTGTGCCCCGACGCGCGGGCGGGGTGTTCTCCGCGTGG 300
Qy 301 agtctcaactgggacctgagtgatggtgtcccgagggggtgtgcgggggcatccgctccgc 360
Db 301 AGTCTCACTGGGACCTGAGTGAATGGCTCCAGGGGCTGTGCGGGGCATCCGCTCCGC 360
Qy 361 cttctcaacagcgtgtgtctcctggaagatgctagcaatggggcgctgcagagat 420
Db 361 CTCTCCACAGCCCTGTGTCTCTGGAAGATGCTAGCAATGGGGCGCTGGCAGGAT 420
Qy 421 tctggatcctctgctcctacttatggttacctgtcctcgtgggcaggccttagaagagg 480
Db 421 TCTGGATCCTCTGCTCTCACTATGTTACCTGTCTCTGGGGCCAGGCTTAGAAGAGG 480
Qy 481 aggaagaaggggccttactagctcaagctgagagaaactagagcccaagcaactcca 540
Db 481 AGGAAGAAGGGGCGCTTACTAGCTCAAGCTGGAGAGAAACTAGAGCCCAAGCAACTCCA 540
Qy 541 cctccagcccatctcatcttctcctagcgatgatcaggagatttagagatgtgggtt 600
Db 541 CCTCCAGCCCATCTCATTTTCTCCTAGCGGATGATCAGGGATTTTAGAGATGTGGGT 600
Qy 601 acccaggtatctgagatataaacacactactcttgacaagctcgtgcggaaggagttaaac 660
Db 601 ACCAGGATCTGAGATTAACACCTACTCTTGACAGCTCGCTGCCGAAGAGGTAAAC 660
Qy 661 tgggaactactatgtccagcctatbtgacacaccatccagagtgatgttattactggaa 720
Db 661 TGGAGAACTACTATGTCCAGCGCTATTTGCACACCATCCAGGAGTCAGTTTATTACTGAA 720
Qy 721 agtatcagatcacacggagctcaacattctatcataagacctaccaccccaacttt 780
Db 721 AGTATCAGATACACCGGACTTCAACATTCTATCATAAGACCTACCCCAACCAACTGTT 780

Qy 781 tacctctggacaatgcccaccctaccctcagaaactgaaagaggttgatattcaacgcata 840
Db 781 TACCTCTGGAACAATGCCACCCCTACCTCAGAACTGAAGSAGGTTGGATATTCAACGCATA 840
Qy 841 tggctggaaaaatggcacttgggttttttttacaaaaaagaatgcatgccccaccagaaagat 900
Db 841 TGGTTCGAAAAATGGCAGCTTGGGTTTTTACAAAAAAGATGCGTCCCAACAGAGGAT 900
Qy 901 ttgataaccttttttgggtccctttttgggaagtgggaattactatcacacacataaattgtg 960
Db 901 TTGATACCTTTTTTGGTTCCTTTTTTGGGAAGTGGGGATTTACTATACACATCAAAATGTG 960
Qy 961 acagtccctgggagtgctggtctatgactgtgtatgaaaacgacaatgctgctgggactatg 1020
Db 961 ACAGTCTCTGGGATGTGTGCTATGACTTGTATGAAAAACGACAATGCTGCCTGGGACTATG 1020
Qy 1021 acaatggcatactactccacacagatgtacactcagagagtcagacaaatcttagtctccc 1080
Db 1021 ACAATGGCATATACTCCACACAGATGTACACTCAGAGAGTACAGCAAAATCTTAGCTTCC 1080
Qy 1081 ataaccacacaaagcctatatttttatattgctctatcaagctgttcatccaccactgc 1140
Db 1081 ATAAACCCACAAAGCCTATATTTTATATATTTGCCCTATCAAGCTGTTTCAATCACCACGTC 1140
Qy 1141 aagctccctggcaggtatttcgaacactaccgatccattatcaacataaacagagagagat 1200
Db 1141 AAGCTCTCTGGCAGGTATTTTCGAACACTACCGATCCATTATCAACATATAACAGAGAGAT 1200
Qy 1201 atgtgcatactcttccctgcttagatgaagcaatacaaaacgtgcattggtctctaaaga 1260
Db 1201 ATGTGCGCATGCTTTCTCTTAGATGAAGCAATCAACAACGTGACATTTGGCTCTTAAAGA 1260
Qy 1261 cttatggtttctataacaacagcattatcttactcttccagataatggtggccagccta 1320
Db 1261 CTTATGGTTTCTATAAACACAGCATATCATTTTACTCTTTCAGATTAATGGTGGCCACCTA 1320
Qy 1321 cggcagggaggagtaactggcctctcagaggttagcaaaaggaaacatatgggaaggagaga 1380
Db 1321 CGCGAGGAGGAGTAACTGGCTCTCAGAGGTAGCAAAAGGAACAATATTGGGAAGAGGGA 1380
Qy 1381 tccgggctgtagggtttgtgatagccacctctgaaaaaaagaaggaacagtgctaaagg 1440
Db 1381 TCCGGGCTGTAGGCTTTGTGCATAGCCCACTTCTGAAAAACAAAGGAACAGTGTGTAAAG 1440
Qy 1441 aacctgtcacatacactgactggttaccacactctatttctgctgctgaagacagatcg 1500
Db 1441 AACTTGTGCACATCACTGACTGCTACCCCACTCTCATTTTCACTGGCTGAAGACAGATTG 1500
Qy 1501 atgaggacattcaactagatggtatgatatctgggagaccataagtgagggtctctcgt 1560
Db 1501 ATGAGGACATTCACATAGATGGCTATGATATCTGGGAGACCATAAGTGAGGCTCTTCGCT 1560
Qy 1561 caccocagtagatatttttgataaacaatgacccccataacaccaaagcgaataatgct 1620
Db 1561 CACCCGAGTAGATATTTTGCATAACATTTAGCCCCATATACACCAAGGCAAAATTTGGCT 1620
Qy 1621 cctgggacagcaggtatggatctggaacactgcaatccagtcagccatcagagtcagac 1680
Db 1621 CCTGGGACAGGCTATGGGATCTGGAACACTGCAATCCAGTCAGCCATCAGAGTGCAGC 1680
Qy 1681 actgaaattgctttacagaaatcctggctacagcactgggtccccccctcagttctta 1740
Db 1681 ACTGAAATTTGCTTACAGGAAATCTGGCTACAGCGACTGGGTCCCCCTCAGCTTTTCA 1740
Qy 1741 gcaacctgggaccgaaccgggtggcaaatgaacgaatcaacctcgtaactggcaaaagt 1800
Db 1741 GCAACTGGGACCGAACCCTGGTGGCAATGAACCGGATCACTTGTCACTGCGCAAAAGTG 1800
Qy 1801 tatgcttttcaacatacacagccagcccatatgagaggttgacccatcttaacaggtatc 1860
Db 1801 TATGCTTTTCAACATACAGCCGACCCCATATGAGAGGGTGGACCTATCTTAACAGGTATC 1860
Qy 1861 caggaaatcgtaagaagctcctacgaggtctctcagttcaacaaaaactcagtgccgg 1920

Db 1321 TGGAAACACTGCAATCCAGTCAGCCCATCAGAGTGCGAGCACTGGAAATTCCTTACAGGAAAT 1380
Qy 1704 cctggctacagcagctgggtccccctcagctctttcagcaaacctgggaccgaacccggtgg 1763
Db 1381 CCTGGCTACAGGACTGGGTCCCCCTCAGTCTTTTCACCAACCTGGGACCGGCTGG 1440
Qy 1764 cacaatgaacggatcacctcgtcaactggcaaaagtgtatggcttttcaacatcacagcc 1823
Db 1441 CACAATGAACGGATCACCTTGTCAACTGGCAAAAGTGATGGCTTTTCAACATCACAGCC 1500
Qy 1824 gaccatatgaggggtggaactatctcaacagggtatccaggaatcgtgaagaagctccta 1883
Db 1501 GACCATATGAGAGGTGGAGCTTATCTAACAGGTATCCAGGAATCGTGAAGAAGCTCCTA 1560
Qy 1884 cggaggtctcacagttcaacaactcagtcggtcggtcaggtatcccccaaaagacccc 1943
Db 1561 CGGAGGCTCTCACAGTTCAACAAACTGCAGTGGCGGTACAGGTATCCCCCAAGAGCCCC 1620
Qy 1944 agaagtaacctaggtcgaatggagggtct 1974
Db 1621 AGAAGTAAACCTAGGCTCAATGGAGGGGTCT 1651

RESULT 3
AC104779/c
LOCUS AC104779 152861 bp DNA linear PRI 09-FEB-2002
DEFINITION Homo sapiens chromosome 4 clone RP11-26P13, complete sequence.
ACCESSION AC104779
VERSION AC104779.4 GI:18642899
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 152861)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 152861)
Waterston,R.H.
Direct Submission
Submitted (21-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 152861)
Waterston,R.H.
Direct Submission
Submitted (09-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 9, 2002 this sequence version replaced gi:18482303.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: H.NH0026P13
----- Location/Qualifiers -----
1.152861
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-26P13"
48870 a 27681 c 28505 g 47805 t

BASE COUNT 48870 a 27681 c 28505 g 47805 t
ORIGIN

Query Match 53.3%; Score 1207; DB 9; Length 152861;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 722 gtatcagatcacacgggacttcaacattctatcataagacactcccccaactggttt 781
Db 10196 GTATCAGATACACACGGACTTCAACATTCATCATAGAGCTACCAACCACTGTTT 10137
Qy 782 acctctggacaatgcccacccctacatcagaactgaaggttgatattcaacgcatat 841
Db 10136 ACCTCTGCACAAATGCCACCTTACCTGCAAAACTGAAGGAGTGGATATTCACGCATAT 10077
Qy 842 ggtcggaaaaatggcaacttgggtttttacagaaaaaagaatgcatgcccacagagaaggatt 901
Db 10076 GGTCCGAAAAATGGCACTTGGGTTTTTACAGAAAAAGAAATGCATGCCACACAGAGAGATT 10017
Qy 902 tgataaccttttgggttcccttttgggaagtgggtattactatcacactacaatgaga 961
Db 10016 TGATACCTTTTGGTTCCTTTTGGGAAGTGGGATTTACTATACACTACAAATGTGA 9957
Qy 962 cagtcctgggagtgtgtgctatgactgtatgaaaaacacaaatgctgctgggactatga 1021
Db 9956 CAGTCTGGGATGTGTGGCTATGACTTGTATGAAAAACGACAATGCTGCTGGGACTATGA 9897
Qy 1022 caatggcatatactccacacagagtacactcagagagtagcagaatacttagcttccca 1081
Db 9896 CAATGGCATATACTCCACACAGATGTACACTCAGAGAGTACAGCAAAATCTTATAGCTTCCCA 9837
Qy 1082 taacccccaaagcctatatattttatatatgtcctatcaagctgttctaccactatga 1141
Db 9836 TAACCCCAACAAGCTATATTTTATATATATGCTTATCAAGCTGTTTATTCACCACTGCA 9777
Qy 1142 agtcctggcagggtatttcgaacactaccgatccatctcaacaataacagagagagata 1201
Db 9776 AGTCTCTGGCAGGTATTTTCGAACACTACCGATCCATTTATCAACATAAACAGGAGGAGATA 9717
Qy 1202 tgcgtccatgttttctcgtcttagatgaagcaaatcaacaacgtgacatggctctaaaagac 1261
Db 9716 TGCTGCCATGCTTCTCCTTAGATGAAGCAATCAACAACCTGACATTTGGCTCTTAAGAC 9657
Qy 1262 ttatggtttctataacaacacagcattatcttactctcagataaatggtggcagccctac 1321
Db 9656 TTATGGTTTCTTAAACACACAGCATTTATCATTTACTTTCAGATATGTTGGCCACCTTAC 9597
Qy 1322 ggcaggaggagtaactggcctctcagaggtagcaaaagaaacatatattgggaaggagat 1381
Db 9596 GGCAGGAGGAGTAACTGGCTCTCAGAGGTAGCAAGGAACATATTTGGGAAGGAGGAT 9537
Qy 1382 ccgggcttaggctttgtgcatagcccaactctctgaaaaacaagggaacagtgatgaaga 1441
Db 9536 CCGGCTGTAGGCTTTGTGCATAGCCCCACTTCTGAAAAACAAGGGAACAGTGTGTGAAGGA 9477
Qy 1442 acctgtgcacatcactgactggtaccacactctctactctgctgaggaagacagatga 1501
Db 9476 ACTTGTGCACATCACTGACTGCTGCTACCCCTCTCTCTTCTGCTGGAAGGACAGATTGA 9417
Qy 1502 tgaggacattcaactagatggctatgatatctgtggagaccataagtggggtcttcgctc 1561
Db 9416 TGAGGACATTCAACTAGATGGCTATGATATCTGGGAGACCATAAGTAGGGGTCTTCGCTC 9357
Qy 1562 acccgaatagatatatttgcaataacattgaccatatacaccaaggcaaaaatggctc 1621
Db 9356 ACCCGAGTAGATATTTTGCATACATTGACCCCATATACACCAAGGCAAAAATGGGCTC 9297
Qy 1622 ctgggagcaggctatgggactctggaacactgcaatccagtcagccatcagagtgcagca 1681
Db 9296 CTGGGACACAGGCTATGGGATCTGGAACACTTGAATTCAGTCAGGCATCAGAGTGCAGCA 9237
Qy 1682 ctggaattgtcttacaggaaatctcgtcagcagactgggtctccccctcagcttctcag 1741
Db 9236 CTGGAATTTGCTTACAGGAAATCTGGCTACAGCGACTGGGTCCCTCCCTCTAGTCTTTCAG 9177
Qy 1742 caactgggagccgaacgggtggcaaatgaacgataccctcgtaacagcaaaaggt 1801
Db 9176 CAACCTGGGCCGGAACCCGGTGGCAATGAACGGATCACTTGTCTCACTTGGCAAAAGTGT 9117

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4777
Center clone name: 24_I_3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156719 bases at least Q40
Consensus quality: 160377 bases at least Q30
Consensus quality: 162195 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 163847; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 867: contig of 867 bp in length
* 868 967: gap of 100 bp

```

```

2234 2233: gap of 100 bp
* 2234 3987: contig of 1754 bp in length
* 3988 4087: gap of 100 bp
* 4088 6971: contig of 2884 bp in length
* 6972 7071: gap of 100 bp
* 7072 10879: contig of 3808 bp in length
* 10880 10979: gap of 100 bp
* 10980 15646: contig of 4667 bp in length
* 15647 15746: gap of 100 bp
* 15747 25208: contig of 9462 bp in length
* 25209 25308: gap of 100 bp
* 25309 33458: contig of 8150 bp in length
* 33459 33558: gap of 100 bp
* 33559 46617: contig of 13059 bp in length
* 46618 46717: gap of 100 bp
* 46718 60565: contig of 13848 bp in length
* 60566 60665: gap of 100 bp
* 60666 92890: contig of 32025 bp in length
* 92891 92790: gap of 100 bp
* 92791 122716: contig of 29926 bp in length
* 122717 122816: gap of 100 bp
* 122817 149492: contig of 26676 bp in length
* 149493 149592: gap of 100 bp
* 149593 165147: contig of 15555 bp in length.
Location/Qualifiers
1. 165147
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-2413"
/clone_lib="RPC1-11 Human Male BAC"
1. 867
/note="assembly_fragment"
clone_end:SP6
vector_side:left
968..2133
/note="assembly_fragment"
2234..3987
/note="assembly_fragment"
4088..6971
/note="assembly_fragment"
7072..10879

```

```
/note="assembly_fragment"
10980..15646
/note="assembly_fragment"
15747..25208
/note="assembly_fragment"
25309..33458
/note="assembly_fragment"
33559..46617
/note="assembly_fragment"
46718..60565
60666..92690
/note="assembly_fragment"
92791..122716
/note="assembly_fragment"
122817..149492
/note="assembly_fragment"
149593..165147
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 52130 a 29805 c 30056 g 51853 t 1303 others
ORIGIN

Query Match 53.3%; Score 1207; DB 2; Length 165147;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 722 gatacagatacacaccggacttcaacattctatcatagaacctaccacccaactgttt 781
Db 101880 GTATCAGATACACACCGGACTTCAACATTCATATAGAGACTACCCCAACCAACTGTTT 101821

Qy 782 acctctggacaatgccaccctacctcagaaactgaagaggttgatatccaacgcatat 841
Db 101820 ACCCTGGACAATGCCACCCTACCTCGAGAACTGAAGAGGTTGGATATTCAACGCGATAT 101761

Qy 842 ggtcggaaaaatggcacttgggtttttacagaaaagaaatgcatagccaccagaaagagatt 901
Db 101760 GGTGCGAAAATGGCACTTGGGTTTTCACAGAAAAGAAATGCATGCCACCACGAAGGATT 101701

Qy 902 tgataccttttttgggttcccttttgggaagtggggattactatatacacactacaatgtga 961
Db 101700 TGATACCTTTTTTGGTTCCCTTTTGGGAAGTGGGGATTACTATACACTACAAATGTGA 101641

Qy 962 cagtcctggagatgtggctatgactgtatgaaaacagacaatgcctcctgggactatga 1021
Db 101640 CAGTCCTGGGATGTGGCTATGACTTGTATGAANAACGACAATGCTGCTGGGACTATGA 101581

Qy 1022 caatggcatatactccacacagatgtacactcagagagtacagcaaatcttagcttccca 1081
Db 101580 CAATGGCATATATCCACACAGATGTACACTCAGAGAGTACAGCAAAATCTTAGCTTCCCA 101521

Qy 1082 taaccccaaaaagcctatattttatatttgccctataaagcttcaactcaccactgca 1141
Db 101520 TAAACCCCAAAAGCCTATATTTTATATATTTGSCCTATATCAAGCTGTTATTCCACACTGCA 101461

Qy 1142 agctcctggcaggatatttcgaacactaccgataccattatcaacaataacagagagagata 1201
Db 101460 AGCTCCCTGGCAGGATTTTCGAANAACCTACCGATCCATTATCAACATTAACAGAGGAGATA 101401

Qy 1202 tgcctgccatgctttcctgcttagtagaagcaatcaacaacgtgacattggctctaaagac 1261
Db 101400 TGCTGCCATGCTTTCCTGCTTAGATGAAGCAATCAACAACGTGACATTGGCTCTAAAGAC 101341

Qy 1262 ttatggtttctataacaacagcatataatcttactcttcagataatgggtgccagcctac 1321
Db 101340 TTATGGTTTCTATAAACACAGCATTTATCTTTACTCTTCAGATATGTTGGTCCAGCCTAC 101281

Qy 1322 ggcaggaggagtaactggcctctcaaggtacagaaaggaacatatiggaagagagat 1381
Db 101280 GGCAGGAGGAGTAACTGGCCCTCTCAGAGGTAGCAAGGAACATATTGGGAAGGAGGAT 101221
```

```
Qy 1382 ccgggctgtaggctttgtgcatagcccaacttctgaaaaacaaggaaacagtggtgaaagga 1441
Db 101220 CCGGGCTGTAGGCTTTGTGCATAGCCCACTTCTGAAAAACAAGGAAACAGTGTGTAAAGGA 101161

Qy 1442 acctgtgcacatcaactgactggttaccocactctctatttcactggtcgtgaagacagatgga 1501
Db 101160 ACTTGTGCACATCACTGACTGGTACCCCACTCTCATTTTCACTGGCTGAAGACAGATTGA 101101

Qy 1502 tgaggaattcaactagatggtctatgatatctggagacacataagtgagggtcttcgcctc 1561
Db 101100 TGAGGACATTCAACTAGATGGCTATGATATCTGGAGACCAATAAGTAGGGGTCTTCGGCTC 101041

Qy 1562 accccgagtagatatatttgcataaaccattgaccccatatacacccaaggcaaaaaatggctc 1621
Db 101040 ACCCCGAGTAGATATTTTGCATAACATTGACCCCATATACACCAAGGCAAAAATGGCTC 100981

Qy 1622 ctggcgacgaggtctatgggatactggaaacactgcaatccagtcagcca tcagagtgacgca 1681
Db 100980 CTGGCGACGAGGCTATGGGATCTGGAACACTGCAATCCAGTCCAGTCAGGCATCAGAGTGCAGCA 100921

Qy 1682 ctggaattgcttacagaaaatcctgctacagcactgggtcccccctcagttcttcag 1741
Db 100920 CTGGAATTGCTTACAGGAAATCCTGGCTACAGCACTGGGTGCCCTCTAGTCTTTTCAG 100861

Qy 1742 caacctgggaccggaaccggtggcacaatgaacgcatcaccctcgtcaactggcaaaaagtgt 1801
Db 100860 CAACCTGGGACCGGAACCGGTGGCACAAATGAACGGATCACTTGTCAACTGGCAAAAAGTGT 100801

Qy 1802 atgcttttcaaatcacagccgacccaatatgagaggggtggacctatctacaaggtatcc 1861
Db 100800 ATGCTTTTCAACATCACAGCCGACCCCATATGAGAGGGTGGACCTATCTTAACAGGTATCC 100741

Qy 1862 aggaatcgtgaagaagctcctacgaggtctcacagttcagttcaacaaaactgcagtcgcggt 1921
Db 100740 AGGAATCTGGAAGAAGCTCTTACGAGGCTCTCACAGTTTCAACAAAACCTGCAGTGCCTGGT 100681

Qy 1922 caggtatcccccccaagaacccccagaagtaaccttaggtcattgaggggtctcaggggacc 1981
Db 100680 CAGGTATCCCCCAAGACCCCAAGAGTAACCTTAGGCTCAATGGAGGGTCTGGGGACC 100621

Qy 1982 atggtatagagagaacccaagaaagaagcgaagcaaaaaatcaggctgagaaaaagca 2041
Db 100620 ATGGTATAAAGAGAAACCAAGAAAGAAAGAACCCAAAGCAAAAATCAGGCTGAGAAAAAGCA 100561

Qy 2042 aagaagaaagc-aaaaaagaagaagaaacagacagaagcagttcaggttcaacttcgca 2100
Db 100560 AAAGAAAAGCAAAAAGAAAAGAAAGAAACACAGCAAGAACAGTCTCAGGTTTCAACTTGCCA 100501

Qy 2101 ttcagggttacttctggtgataagcacaaaataattcctgtttgtgttaaaactttaatcagtt 2160
Db 100500 TTCAGGTGTTACTTGTGGATAAGCACAAATATTTCCCTGTTTGGTTAAACTTTAATCACTT 100441

Qy 2161 cttatcttcatctgtttcctcaggttaaacacagcaaaatttggctcgataataatcgtcgacc 2220
Db 100440 CTTATCTTTCATCTGTCTTCCTAGGTAAACACAGCAAAATTTGGCTCGATAATATATCGTGGCC 100381

Qy 2221 taagcgtcaggctgttttctcagtcgtgcccac 2252
Db 100380 TAAGCGTCAGGCTGTTTTTCATGCTGTGCCAC 100349
```

RESULT 5

AC013692/c

LOCUS

DEFINITION

AC013692

AC013692

AC013692

AC013692

AC013692

AC013692

AC013692

AC013692

AC013692

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45

Homo sapiens clone RP11-21L

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

1 (bases 1 to 157043)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome, clone RP11-2118
 Unpublished
 2 (bases 1 to 157043)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,K., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehotzky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (13-NOV-1999), Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 7, 2000 this sequence version replaced gi:10280848.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4053
 Center clone name: 21_L_8

TITLE JOURNAL COMMENT

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 45 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 924: contig of 924 bp in length
 * 925 1024: gap of 100 bp
 * 1025 1800: contig of 776 bp in length
 * 1801 1900: gap of 100 bp
 * 1901 2800: contig of 900 bp in length
 * 2801 2900: gap of 100 bp
 * 2901 3628: contig of 728 bp in length
 * 3629 3728: gap of 100 bp
 * 3729 4582: contig of 854 bp in length
 * 4583 4682: gap of 100 bp
 * 4683 5763: contig of 1081 bp in length
 * 5764 5863: gap of 100 bp
 * 5864 6992: contig of 1129 bp in length
 * 6993 7092: gap of 100 bp
 * 7093 8691: contig of 1599 bp in length
 * 8692 8791: gap of 100 bp
 * 8792 10237: contig of 1446 bp in length
 * 10238 10337: gap of 100 bp
 * 10338 12115: contig of 1778 bp in length
 * 12116 12215: gap of 100 bp
 * 12216 14480: contig of 2265 bp in length
 * 14481 14580: gap of 100 bp
 * 14581 16200: contig of 1620 bp in length
 * 16201 16300: gap of 100 bp
 * 16301 17801: contig of 1501 bp in length
 * 17802 17901: gap of 100 bp
 * 17902 19560: contig of 1659 bp in length
 * 19561 19660: gap of 100 bp
 * 19661 21061: contig of 1401 bp in length
 * 21062 21161: gap of 100 bp

* 21162 22798: contig of 1637 bp in length
 * 22799 22898: gap of 100 bp
 * 22899 25224: contig of 2326 bp in length
 * 25225 25324: gap of 100 bp
 * 25325 27369: contig of 2045 bp in length
 * 27370 27469: gap of 100 bp
 * 27470 30262: contig of 2793 bp in length
 * 30263 30362: gap of 100 bp
 * 30363 32450: contig of 2088 bp in length
 * 32451 32550: gap of 100 bp
 * 32551 34415: contig of 1865 bp in length
 * 34416 34515: gap of 100 bp
 * 34516 37321: contig of 2806 bp in length
 * 37322 37421: gap of 100 bp
 * 37422 40476: contig of 3055 bp in length
 * 40477 40576: gap of 100 bp
 * 40577 50235: contig of 9659 bp in length
 * 50236 50335: gap of 100 bp
 * 50336 53457: contig of 3122 bp in length
 * 53458 53557: gap of 100 bp
 * 53558 56770: contig of 3213 bp in length
 * 56771 56870: gap of 100 bp
 * 56871 59626: contig of 2756 bp in length
 * 59627 59726: gap of 100 bp
 * 59727 62247: contig of 2521 bp in length
 * 62248 62347: gap of 100 bp
 * 62348 65709: contig of 3362 bp in length
 * 65710 65809: gap of 100 bp
 * 65810 69014: contig of 3205 bp in length
 * 69015 69114: gap of 100 bp
 * 69115 73176: contig of 4062 bp in length
 * 73177 73276: gap of 100 bp
 * 73277 78978: contig of 5702 bp in length
 * 78979 79078: gap of 100 bp
 * 79079 83397: contig of 4319 bp in length
 * 83398 83497: gap of 100 bp
 * 83498 87887: contig of 4390 bp in length
 * 87888 87987: gap of 100 bp
 * 87988 91530: contig of 3543 bp in length
 * 91531 91630: gap of 100 bp
 * 91631 97197: contig of 5467 bp in length
 * 97098 97197: gap of 100 bp
 * 97198 103021: contig of 5824 bp in length
 * 103022 103121: gap of 100 bp
 * 103122 108656: contig of 5535 bp in length
 * 108657 108756: gap of 100 bp
 * 108757 113320: contig of 4564 bp in length
 * 113321 113420: gap of 100 bp
 * 113421 120645: contig of 7225 bp in length
 * 120646 120745: gap of 100 bp
 * 120746 130194: contig of 9449 bp in length
 * 130195 130294: gap of 100 bp
 * 130295 137856: contig of 7562 bp in length
 * 137857 137956: gap of 100 bp
 * 137957 147125: contig of 9169 bp in length
 * 147126 147225: gap of 100 bp
 * 147226 154966: contig of 7741 bp in length
 * 154967 155066: gap of 100 bp
 * 155067 157043: contig of 1977 bp in length.
 FEATURES
 source
 1. .157043
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-2118"
 /clone_l1b="RPC1-11 Human Male BAC"
 1. .924
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 misc_feature
 1025..1800
 /note="assembly_fragment"
 1901..2800
 /note="assembly_fragment"
 misc_feature

DEFINITION Homo sapiens chromosome 4 clone RP11-219L11, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
ACCESSION AC105414
VERSION AC105414.4 GI:186642912
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 175645)
Waterston, R.H.
Direct Submission
Submitted (04-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 9, 2002 this sequence version replaced gi:18483567.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: H.NH0219L11
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 174714 bases at least Q40
Consensus quality: 174877 bases at least Q30
Consensus quality: 174989 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 175545; sum-of-contigs
Quality coverage: 10.91 in Q20 bases; agarose-fp
Quality coverage: 8.85 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 74211: contig of 74211 bp in length
* 74212 74312: gap of unknown length
* 74312 175645: contig of 101334 bp in length.
FEATURES
source Location/Qualifiers
1..175645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-219L11"
misc_feature 1..74211
/note="assembly_name:Contig34"
misc_feature 74312..175645
/note="assembly_name:Contig35"
BASE COUNT 54056 a 32068 c 32415 g 56997 t 109 others
ORIGIN
Query Match 27.8%; Score 630; DB 2; Length 175645;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 680; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 44 ttttttttcttggtggaagctgcttagggagggggaggaggaagaagtga 103

Db 91286 TTTTCTTTTCTCTGGTGGAAAGCTGCTTAGGGAGGGGGAGGAGAGAGTGA 91227
Qy 104 tgtctggagaagagagagccctctgttcttcggaagtcctccattcaagc 163
Db 91226 TGTGCTGGAGAAGAGCGAGCCCTCTCTTCTTCGCGAGTCCATCCATTAAGCCATCAC 91167
Qy 164 ttctggaagattaaagtgtcgacatggtgacagctgagagagagagattcttgc 223
Db 91166 TTCTGGAAGATTAAAGTTGTGCGACATGTCACAGCTGAGAGGAGAGGATTTCTTGC 91107
Qy 224 caggtgagagacttccaccgtctgttgggtgcatgtgtgcgcccgcascgcgggcg 283
Db 91106 CAGGTGGAGAGTCTTACCGCTGTGTTGGTGCATGTGTGCGCCCGCAGCGCGGGCG 91047
Qy 284 cgtggtctccgctggaagtctccctggaccctgagtggaatgctccaggggctg 343
Db 91046 CGTGGTCTCCGCGTGGAGTCTCACCTGGGACCTGAGTGAATGGCTCCAGGGGCTGTGC 90987
Qy 344 ggggcacgcctccgcttctccacagggcgtgtgtctgtctctggaagatgctagcaat 403
Db 90986 GGGCATCCGCTCCGCCCTTCTCCACAGSCCTGTGTCTCTGGAAGATGCTAGCAAT 90927
Qy 404 gggggcgtggcaggattctggtctctgctcctcctcattgttaccctgtcctgggg 463
Db 90926 GGGGCGCTGGCAGGATTTCTGATCTCTGCTCTCTCACTTATGTTACCTGTCTCTGGGG 90867
Qy 464 ccaggccttagaagagagaagaaggggccttactagctcaagctgagagaactaga 523
Db 90866 CCAGGCTTAAAGAGAGGAGGAAGAGGGGCTTACTAGCTCAAGCTGAGAGAACTAGA 90807
Qy 524 gccagacaacttccacctccagcccatctcatctttctcctagcgatgcaggg 583
Db 90806 GCCAGCACACTTCCACCTCCAGCCCATCTCATTTTCTCTAGCGGATGATCAGG 90747
Qy 584 atttagagatggtgtaccacgagatctgagataaaacacctactcttgacaagctgc 643
Db 90746 ATTTAGAGATGTTGGTTTACCACGGATCTGAGATTAAACACACTACTCTTGACAAGCTGCC 90687
Qy 644 tgcgaagagattaaactggaactactatgtccagctattgtccagctattgtccacacatccagag 703
Db 90686 TGCCGAAGAGTTAACTGGAGAACTACTATGTCCAGGCTATTTTCACACCATCCAGGAG 90627
Qy 704 tcagttattactggaagta 724
Db 90626 TCAGTTATTACTGGAAAGTA 90606
RESULT 8
AK027201 1871 bp mRNA linear PRI 29-SEP-2000
LOCUS Homo sapiens CDNA: FLJ23548 fis, clone LNG08487.
DEFINITION AK027201
ACCESSION AK027201
VERSION AK027201.1 GI:10440271
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens human lung cDNA to mRNA, clone LNG08487.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,
Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Tanigami, A.,
Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiroo, M.,
Ohmori, Y., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T.,
Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 1871)
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human

Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES
source Location/Qualifiers
i. 1871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="LNG08487"
/tissue_type="human lung"
/clone_lib="LNG"
/note="cloning vector pME18SFL3"
204..713
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15689.1"
/db_xref="GI:10440272"
/translation="MAPQQWMSGTLQSSQSECTGRLLTGNPGYSDWVPQSFNS
LGNRNRERITLSTGKSWLNFITADPYERVDLSNRPVIGIVKLLRLSOFNKTAVP
VYPPKDRPSNRLNGVWGPWYKEETKKKPKSKNOAEKKQKKKKKKQKRAVSGS
TCHSGVTCG"
BASE COUNT 604 a 367 c 354 g 546 t
ORIGIN

Query Match 11.6%; Score 263; DB 9; Length 1871;
Best Local Similarity 99.1%; Pred. No. 9.3e-143;
Matches 753; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
QY 1495 agattgatgagacattcaactagatggtctatgatctgtggagaccataagttagggtc 1554
|||||
Db 84 AGATTGATGAGGACATTCAACTAGATGGCTATGATCTGGGAGACCATAAGTGAGGGTC 143
QY 1555 ttgcgtcaccccgagtagattttgcataacattgaccccatatacaccaagcagcaaaa 1614
|||||
Db 144 TTCGCTCACCCGAGTAGATATTTTGTGATACATTGACCCCATATACACCAAGGCAAAA 203
QY 1615 atggtcctctggcagcaggctatgggtctgtggaacactgcaatcagtcagccatcagag 1674
|||||
Db 204 ATGGCTCTGGCAGCAGGCTATGGGATCTGGAACACTTGCAATCCAGTCAGCCATCAGAG 263
QY 1675 tgcagcact-ggaattgcttaaggaaatccttggtctacagcagctgggtccccctcag 1733
|||||
Db 264 TGCAGCACTGGGAGATTGCTTTACAGGAAATCCTGGCTACACGACTGGGTCCCCCTCAG 323
QY 1734 tctttcagcaacctgggacccaacccggtggcacaatgaacggatcacctcgtcaactggc 1793
|||||
Db 324 TCTTTCAGCACTTGGACCGAACCGGTGGCGCAATGACCGATCACCTTGTCAACTGGC 383
QY 1794 aaaagtgtatggcttttcaacatcacagccagccatataatgaggggtggaacctatcaac 1853
|||||
Db 384 AAAAGTGTATGGCTTTTCAACATCACAGCCGACCATATGATGAGGGTGGACCTATCTAAC 443
QY 1854 aggtatccaggaalcgtagaagaactcctcagggaggtctctcacagttcaacaaactgca 1913
|||||
Db 444 AGGTATCCAGGAATCGTGAAGAAGCTCCTACGGAGGCTCTCACAGTTTCAACAAACTGCA 503
QY 1914 gtccggtccaggtatcccccaagaccccaagaaagtaacctaggctcaatgaggaggtc 1973
|||||
Db 504 GTGCCGTCAGGTATCCCCCAAGACCCCAAGAGTAACTAGCTCAATGGAGGGGTC 563
QY 1974 tagggacatgtgtatagagggaaaccaaagaaaagcaagccaaagcaaaatcaggctgag 2033
|||||
Db 564 TGGGACCATGTTAAGAGAGAAACCAAGAAAAGAGCCCAAGACAAATCAGGCTGAG 623
QY 2034 aaaaagcaaaagaaagc-aaaaaagaagaagaaacagcagaagcagtcagtcagggtca 2092
|||||

Db 624 AAAAAAGCAAAAGAAAAAGCAAAAGAAAAAGAAAGAAACAGCAAGCAAGCTCTCAGGTTCA 683
QY 2093 acttgccattcaggtgttaacttggtgataagcacaataatttcctgttggttaaacttt 2152
|||||
Db 684 ACTTGCCCAATTGAGGTGTACCTGTGGATAAGCACAATAATTTCCTGTGGTTAAACATT 743
QY 2153 aatcagttcttcttcttctatctgttcttctagtagtaaacaccagcaaaatttggtcgcataaat 2212
|||||
Db 744 AATCAGTCTTATCTTTCATCTGTTTCTAGTAGTAACCAAGCAAAATTTGGCTCGATAATAT 803
QY 2213 cgctggcctaagcgtcaggctgttttctcatgctgtgccac 2252
|||||
Db 804 CGCTGGCCTAAGCGTCAGGCTTGTTCATGCTGTGCCAC 843
RESULT 9
AC105383
LOCUS
DEFINITION Homo sapiens chromosome 4 clone RP11-9G1, WORKING DRAFT SEQUENCE, 4
unordered pieces.
AC105383 AC031992
VERSION AC105383.2 GI:18376928
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Plimates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 176341)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176341)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jan 26, 2002 this sequence version replaced gi:18042331.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_NH0009G01
Drafting center: WIBR
----- Summary Statistics -----
Sequencing vector: M13; 34%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 173694 bases at least Q40
Consensus quality: 174243 bases at least Q30
Consensus quality: 174882 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 176041; sum-of-contigs
Quality coverage: 11.08 in Q20 bases; agarose-fp
Quality coverage: 10.78 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1053: contig of 1053 bp in length
* 1054 1153: gap of unknown length
* 1154 5581: contig of 4428 bp in length

* 5582	5681: gap of unknown length
* 5682	63724: contig of 58043 bp in length
* 63725	63824: gap of unknown length
* 63825	176341: contig of 112517 bp in length.
FEATURES	Location/Qualifiers
source	1..176341
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="4"
	/clone="RP11-9G1"
misc_feature	1...1053
	/note="assembly_name:Contig33"
misc_feature	1154..5581
	/note="assembly_name:Contig37
	clone_end:SP6
	vector_side:left
misc_feature	5682..63724
	/note="assembly_name:Contig38
	clone_end:T7
	vector_side:left
misc_feature	63825..176341
	/note="assembly_name:Contig39"
BASE COUNT	61266 a 30790 c 29844 g 54139 t 302 others
ORIGIN	
Query Match	4.4%; Score 99; DB 2; Length 176341;
Best Local Similarity	99.0%; Pred.No. 5.1e-46;
Matches 199; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Qy	2052 aaaaaagaagaaacagcagcaaacgagtcttcaggttccaacttgccattgcagggttta 2111
Db	97 AAAAAAGAGAAGAACACAGCAAGACAGCTTCAGGTTCAACTTGCATTCAGTGTTA 156
Qy	2112 ctctggatgaagcacacaatatcttctctgttgtaacttaataatcagttcttatcttcaa 2171
Db	157 CTCTGGATAAGCACAATAATTCTCTGTTGGTTAAACTTTTAATCAGATCTTATCTTTCA 216
Qy	2172 tctgtttccttagttaaaccagcaaatctgctcgataataatcgctgcctaagcgtcagg 2231
Db	217 TCTGTTTCCTAGGTAACACAGCAAATTTGGCTCGATAATATFCGTGCCCTAAGCGTCAGG 276
Qy	2232 ctgtttttcatgctgtgccac 2252
Db	277 CTTGTTTTTCATGCTGCTGCCAC 297
RESULT 10	
AC091322/c	
LOCUS	AC091322 232951 bp DNA linear HTG 11-JUN-2001
DEFINITION	Mus musculus clone RP23-60D3, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION	AC091322
VERSION	AC091322.2 GI:14336533
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	house mouse.
ORGANISM	Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
1 (bases 1 to 232951)	
Birren,B., Linton,L., Nusbaum,C. and Lander,E.	
Mus musculus, clone RP23-60D3	
Unpublished	
2 (bases 1 to 232951)	
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,	
Barna,N., Bastien,V., Boguslavsky,I., Boukhgalter,B., Brown,A.,	
Camarat,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,	
Collins,S., Collamore,A., Cooke,P., DeArellano,K., Dewar,K.,	
Diaz,J.S., Dodge,S., Fero,S., Ferreira,P., Fitzhugh,W., Gage,D.,	
Galanen,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,	
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,	
Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,	
Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G.,	
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,	
McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Menesius,L.,	
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,	
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,	
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Roman,C.,	
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,	
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,	
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,	
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,	
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,	
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,	
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.	
Direct Submission	
Submitted (15-APR-2001) Whitehead Institute/MIT Center for Genome	
Research, 320 Charles Street, Cambridge, MA 02141, USA	
On Jun 11, 2001 this sequence version replaced gi:13625486.	
All repeats were identified using RepeatMasker:	
Smit, A.F.A. & Green, P. (1996-1997)	
http://ftp.genome.washington.edu/RM/RepeatMasker.html	
----- Genome Center	
Center: Whitehead Institute/ MIT Center for Genome Research	
Center code: WIBR	
Web site: http://www-seq.wi.mit.edu	
Contact: sequence_submissions@genome.wi.mit.edu	
----- Project Information	
Center project name: L13252	
Center clone name: 60_D_3	
----- Summary Statistics	
Sequencing vector: Plasmid; n/a; 100% of reads	
Chemistry: Dye-terminator Big Dye; 100% of reads	
Assembly program: Phrap; version 0.960731	
Consensus quality: 231176 bases at least Q40	
Consensus quality: 231965 bases at least Q30	
Consensus quality: 232319 bases at least Q20	
Insert size: 210000; agarose-fp	
Insert size: 232551; sum-of-contigs	
Quality coverage: 10.7 in Q20 bases; agarose-fp	
Quality coverage: 9.6 in Q20 bases; sum-of-contigs	

* NOTE: This is a 'working draft' sequence. It currently	
* consists of 5 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
*	1 53: contig of 53 bp in length
*	54 153: gap of 100 bp
*	154 2590: contig of 2437 bp in length
*	2591 2690: gap of 100 bp
*	2891 11774: contig of 9084 bp in length
*	11775 11874: gap of 100 bp
*	11875 100524: contig of 88650 bp in length
*	100525 106524: gap of 100 bp
*	100625 232951: contig of 132327 bp in length.
FEATURES	Location/Qualifiers
Source	1..232951
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/clone="RP23-60D3"
	/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature	1..53
	/note="assembly_fragment
	clone_end:SP6
	vector_side:left
misc_feature	154..2590
	/note="assembly_fragment"
misc_feature	2691..11774
	/note="assembly_fragment"
misc_feature	11875..100524
	/note="assembly_fragment"
misc_feature	100625..232951
	/note="assembly_fragment"

TITLE
JOURNAL

COMMENT

/note="assembly_fragment
clone_end:17
vector_side:right"
BASE COUNT 70795 a 45474 c 46842 g 69434 t 406 others
ORIGIN

Query Match 2.3%; Score 53; DB 2; Length 232951;
Best Local Similarity 100.0%; Pred. No. 7e-19;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1611 aaaaatgctctctggcagcgatggtggtatctgggaactgcgaatccagtc 1663
|||||
Db 146855 AAAAAAGCTCTGGCGAGCGATGATGGATCTGGACACATGCAATCCAGTC 146803

RESULT 11
AC098604
LOCUS
DEFINITION
AC098604 122992 bp DNA linear HTG 20-DEC-2001
***, 50 unordered pieces.
AC098604
VERSION
AC098604.4 GI:17973771
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 122992)
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alisbrooks,S.L., Amaral, H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,K.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
Ogah,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Thomas,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlecyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
2 (bases 1 to 122992)
AUTHORS
Worley,K.C.

Direct Submission
Submitted (26-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064604.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKIS
Center clone name: CH230-139G21
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 115001 bases at least Q40
Consensus quality: 121003 bases at least Q30
Consensus quality: 125541 bases at least Q20
Estimated insert size: 103320; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 5288: contig of 5288 bp in length
* 5289: gap of unknown length
* 5389: contig of 7434 bp in length
* 12822: gap of unknown length
* 12923: contig of 5998 bp in length
* 18921: gap of unknown length
* 19021: contig of 5251 bp in length
* 24271: gap of unknown length
* 24272: contig of 4721 bp in length
* 29092: gap of unknown length
* 29192: contig of 4128 bp in length
* 33320: gap of unknown length
* 33420: contig of 4814 bp in length
* 38234: gap of unknown length
* 38235: contig of 3138 bp in length
* 38335: gap of unknown length
* 41473: contig of 3184 bp in length
* 41573: gap of unknown length
* 44757: contig of 4648 bp in length
* 44857: gap of unknown length
* 49504: contig of 2163 bp in length
* 49505: gap of unknown length
* 49604: contig of 1622 bp in length
* 51767: gap of unknown length
* 51867: contig of 1356 bp in length
* 53490: gap of unknown length
* 53589: contig of 1920 bp in length
* 54945: gap of unknown length
* 54946: contig of 3564 bp in length
* 55046: gap of unknown length
* 56963: contig of 2686 bp in length
* 56965: gap of unknown length
* 60629: contig of 1727 bp in length
* 60729: gap of unknown length
* 63415: contig of 1672 bp in length
* 63416: gap of unknown length
* 63515: contig of 1411 bp in length
* 65242: gap of unknown length
* 65342: contig of 1202 bp in length
* 65343: gap of unknown length
* 67015: contig of 1672 bp in length
* 67114: gap of unknown length
* 68525: contig of 1411 bp in length
* 68526: gap of unknown length
* 68626: contig of 2102 bp in length
* 70727: gap of unknown length
* 70827: contig of 2102 bp in length
* 70728: gap of unknown length

* 70828 73751: contig of 2924 bp in length
* 73752 73851: gap of unknown length
* 73852 76742: contig of 2891 bp in length
* 76743 76842: gap of unknown length
* 76843 76883: contig of 1841 bp in length
* 76883 78784: gap of unknown length
* 78784 80648: contig of 1865 bp in length
* 80648 80748: gap of unknown length
* 80748 81958: contig of 1210 bp in length
* 81958 82058: gap of unknown length
* 82058 83786: contig of 1728 bp in length
* 83786 83887: gap of unknown length
* 83887 85421: contig of 1535 bp in length
* 85421 85522: gap of unknown length
* 85522 87316: contig of 1794 bp in length
* 87316 89551: gap of unknown length
* 89551 89652: gap of unknown length
* 89652 91451: contig of 1699 bp in length
* 91451 91550: gap of unknown length
* 91550 93240: contig of 1690 bp in length
* 93240 93341: gap of unknown length
* 93341 95063: contig of 1725 bp in length
* 95063 95165: gap of unknown length
* 95165 96781: contig of 1615 bp in length
* 96781 96880: gap of unknown length
* 96880 98569: contig of 1589 bp in length
* 98569 99694: gap of unknown length
* 99694 101631: contig of 1837 bp in length
* 101631 101732: gap of unknown length
* 101732 103235: contig of 1504 bp in length
* 103235 103336: gap of unknown length
* 103336 104663: contig of 1328 bp in length
* 104663 104764: gap of unknown length
* 104764 107050: contig of 2287 bp in length
* 107050 107150: gap of unknown length
* 107150 108692: contig of 1542 bp in length
* 108692 108792: gap of unknown length
* 108792 109912: contig of 1120 bp in length
* 109912 110012: gap of unknown length
* 110012 111516: contig of 1504 bp in length
* 111516 111616: gap of unknown length
* 111616 113034: contig of 1418 bp in length
* 113034 113134: gap of unknown length
* 113134 114537: contig of 1403 bp in length
* 114537 114637: gap of unknown length
* 114637 116034: contig of 1397 bp in length
* 116034 116134: gap of unknown length
* 116134 117734: contig of 1600 bp in length
* 117734 117834: gap of unknown length
* 117834 119020: contig of 1186 bp in length
* 119020 119121: gap of unknown length
* 119121 120235: contig of 1115 bp in length
* 120235 120336: gap of unknown length
* 120336 121616: contig of 1281 bp in length
* 121616 121716: gap of unknown length
* 121716 122992: contig of 1276 bp in length.

FEATURES
source
1..122992
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-139G21"

BASE COUNT 35212 a 23990 c 23255 g 35545 t 4990 others
ORIGIN

Query Match 1.3%; Score 29; DB 2; Length 122992;
Best Local Similarity 100.0%; Pred. No. 0.0001;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 663 gagaaactactatgtccagccatttgcac 691

||||| GAGAACTACTATGTCACCCCTATTTCAC 119505
Db 119477
RESULT 12
BC009528
LOCUS Homo sapiens, clone IMAGE:3890809, mRNA, partial cds.
DEFINITION BC009528 927 bp mRNA linear PRI 22-OCT-2001
ACCESSION BC009528
VERSION BC009528.1 GI:16306920
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 927)
Direct Submission
AUTHORS Strausberg, R.
TITLE CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
JOURNAL DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/BTP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 14 Row: e Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source
1..927
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3890809"
/tissue_type="Lung, carcinoma, large cell
undifferentiated."
/clone_lib="NIH-MGC_69"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
<1..194
/codon_start=3
/product="unknown (protein for IMAGE:3890809)"
/protein_id="AAH09528.1"
/db_xref="GI:16306921"
/translation="DASAHAHASVADIPGPHAAWQISEKWFNSCGAVCLISPGSL
PLTSKGSLSCLRLVSYMGL"

BASE COUNT 267 a 189 c 189 g 282 t
ORIGIN

Query Match 1.1%; Score 26; DB 9; Length 927;
Best Local Similarity 100.0%; Pred. No. 0.0069;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cagcgctccgccacgcgtccgtgga 26
|||||
Db 15 CACGGCTCCGCCACGCGTCCGTGGA 40

RESULT 13
BC012659

LOCUS BC012659 2003 bp mRNA linear ROD 20-AUG-2001
 DEFINITION Mus musculus, Similar to COP9 (constitutive photomorphogenic), subunit 7b (Arabidopsis), clone MGC:13817 IMAGE:4017012, mRNA, complete cds.

ACCESSION BC012659
 VERSION BC012659.1 GI:15215084
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Strausberg, R.

REFERENCE 1 (bases 1 to 2003)
 AUTHORS Direct Submission
 TITLE Submitted (15-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NTH-MGC Project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAK Plate: 18 Row: f Column: 6
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 3309175.

FEATURES
 Location/Qualifiers
 1..2003
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="MGC:13817 IMAGE:4017012"
 /tissue_types="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMV."*
 /clone_11b="NCI_CGAP_Lu29"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 117..911
 /codon_start=1
 /product="Similar to COP9 (constitutive photomorphogenic), subunit 7b (Arabidopsis)"
 /protein_id="AAH12659.1"
 /db_xref="GI:15215085"
 /translation="MAGEQKPSNLLPQFILLAKGTSGSALTTLISQVLEAPGVYVFG ELLELANVQELAGANAYLQLNLFAYGYTPDYIANKEISPELSAQQKLHLIV SLARMKIPYVLLKDLKMRNLELLEIIAVYTDIIQGLKDRQLNLEVDFCIGR DIRKKDINNVTKTLEWDCGEAVLEIQQVLEIRANQYKENHRTQOQVEAYSNIKK TLKATASSAQAEQQLAERECPHPHQPTKQKMSKVGLVSSRH"

BASE COUNT 465 a 541 c 516 g 481 t
 ORIGIN

Query Match 1.1%, Score 26; DB 10; Length 2003;
 Best Local Similarity 100.0%; Pred. No. 0.0067;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caccgcgtccgccacgcgtccgtgga 26
 |||||||||||||||||||||||||
 Db 2 CACGCGTCCGCCACGCGTCCGTGGA 27

RESULT 14
 AC108339/c
 LOCUS AC108339 28735 bp DNA linear HTG 27-JAN-2002

DEFINITION Rattus norvegicus clone CH230-109H4, *** SEQUENCING IN PROGRESS
 ACCESSION AC108339
 VERSION AC108339.1 GI:18377119
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 28735)
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayale, M., Banks, T., Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozadó, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., Mcleod, M.P., Meador, M., Mei, G., Metzker, M., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 28735)
 Worley, K.C.

Direct Submission
 Submitted (27-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: GPM2
 Center clone name: CH230-109H4
 ----- Summary Statistics
 Sequencing vector: Plasmid; M77789
 Chemistry: Dye-terminator Big Dye; 98% of reads
 Assembly program: Phrap; version 0.990329
 findPhrapList
 Consensus quality: 14825 bases at least Q40

Consensus quality: 18227 bases at least Q30
Consensus quality: 20289 bases at least Q20
Estimated insert size: 9365; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
 consists of 26 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 884: contig of 884 bp in length
* 885 984: gap of unknown length
* 985 1846: contig of 862 bp in length
* 1847 1946: gap of unknown length
* 1947 2802: contig of 856 bp in length
* 2803 3685: contig of 783 bp in length
* 3686 3786: gap of unknown length
* 3786 4764: contig of 979 bp in length
* 4765 4864: gap of unknown length
* 4865 5922: contig of 1058 bp in length
* 5923 6022: gap of unknown length
* 6023 6936: contig of 914 bp in length
* 6937 7037: gap of unknown length
* 7037 7918: contig of 882 bp in length
* 7919 8019: gap of unknown length
* 8019 8476: contig of 457 bp in length
* 8476 9397: gap of unknown length
* 9397 10370: contig of 874 bp in length
* 10371 10470: gap of unknown length
* 10471 11498: contig of 1028 bp in length
* 11499 11599: gap of unknown length
* 11599 12507: contig of 908 bp in length
* 12507 13015: contig of 409 bp in length
* 13016 13115: gap of unknown length
* 13116 13933: contig of 818 bp in length
* 13934 14033: gap of unknown length
* 14034 14862: contig of 828 bp in length
* 14862 14961: gap of unknown length
* 14961 16124: contig of 1163 bp in length
* 16125 16225: gap of unknown length
* 16225 16914: contig of 690 bp in length
* 16915 17014: gap of unknown length
* 17015 18138: contig of 1124 bp in length
* 18139 18238: gap of unknown length
* 18239 19416: contig of 1177 bp in length
* 19416 19515: gap of unknown length
* 19516 21083: contig of 1568 bp in length
* 21084 21184: gap of unknown length
* 21184 22939: contig of 1756 bp in length
* 22940 23039: gap of unknown length
* 23040 23810: contig of 771 bp in length
* 23811 23910: gap of unknown length
* 23911 25233: contig of 1323 bp in length
* 25234 25334: gap of unknown length
* 25334 27011: contig of 1678 bp in length
* 27012 27112: gap of unknown length
* 27112 28735: contig of 1624 bp in length.

FEATURES
source
1. .28735
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-109H4"

BASE COUNT
8716 a 4798 c 5034 g 7614 t 2573 others

ORIGIN

Query Match 1.1%; Score 25; DB 2; Length 28735;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 ttaacttttttttttttttttttc 55
Db 26740 TTAACCTTTTCTTTTTCCTTC 26716

RESULT 15

CBRG36E21/c CBRG36E21 38866 bp DNA linear INV 04-NOV-2000

LOCUS Caenorhabditis briggsae cosmid G36E21, complete sequence.
DEFINITION AC084563
ACCESSION AC084563.1 GI:11095013
VERSION HTG.
KEYWORDS SOURCE

ORGANISM Caenorhabditis briggsae.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 38866)
AUTHORS Washington University Genome Sequencing Center.
TITLE The C. briggsae Genome Sequencing Project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 38866)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL

COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsplie@watson.wustl.edu

FEATURES
source

1. 38866
/organism="Caenorhabditis briggsae"
/strain="GujArat G16"
/db_xref="taxon:6238"
/clone="G36E21"

BASE COUNT 12988 a 6629 c 7483 g 11766 t
ORIGIN

Query Match 1.1%; Score 25; DB 3; Length 38866;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2043 aagaaaagcaaaaaaagaagaaga 2067
Db 9383 AAGAAAAGCAAAAAAAGAAGAA 9359

Search completed: July 23, 2002, 19:09:16
Job time: 14232 sec

ed Jul 24 10:05:49 2002

1789	tagatgaagcaatcaacacgtgacattggtctaaagacttatggtttctataacaaca	1848	XX	Homo sapiens.
1282	gcattatcttactcttcagataatggtgcccagctctacgagggagtaactggc	1341	XX	OS
1849	gcattatcttactcttcagataatggtgcccagctctacgagggagtaactggc	1908	XX	WO9946281-A2.
1342	ctctcagagtagcaaaagacatattggaaggaggtccgggctgtaggctttgtgc	1968	PD	16-SEP-1999.
1909	ctctcagagtagcaaaagacatattggaaggaggtccgggctgtaggctttgtgc	2028	PF	08-MAR-1999;
1402	atagcccaactctgaaaaaaggaagacagtggtgaaggaggtccgggctgtaggctttgtgc	1461	XX	10-MAR-1998;
1969	atagcccaactctgaaaaaaggaagacagtggtgaaggaggtccgggctgtaggctttgtgc	1521	PR	98US-0077450.
1462	ggtaccacactctcttctcactggtctgaagacagattgatgagacattcaactagatg	2088	PR	98US-0077632.
2029	ggtaccacactctcttctcactggtctgaagacagattgatgagacattcaactagatg	2148	PR	98US-0077641.
1522	gctatgatctggagacacataaagtgtcttcgctcccccagtagatattttgc	1581	PR	98US-0077649.
2089	gctatgatctggagacacataaagtgtcttcgctcccccagtagatattttgc	2148	PR	98US-0077791.
1582	ataacattgacccatatacaccagggcaaaaatggtctcctggtgagcaggtctatggga	1641	PR	98US-0078004.
2149	ataacattgacccatatacaccagggcaaaaatggtctcctggtgagcaggtctatggga	2207	PR	98US-0040220.
1642	tctggaacactgcaatccagtcagcagctcagagtgagcagctggaattgtttacagaa	2267	PR	98US-0078866.
2208	tctggaacactgcaatccagtcagcagctcagagtgagcagctggaattgtttacagaa	1761	PR	98US-0078910.
1702	atcctgggtacagagcactggtcccccctcagttcttcagcaacctgggacccagcgt	2327	PR	98US-0078936.
2268	atcctgggtacagagcactggtcccccctcagttcttcagcaacctgggacccagcgt	1821	PR	98US-0078939.
1762	ggcacaatgaacggatcacctcgtcaactggtgcaaaagtgtatggttttcaacatcacag	2387	PR	98US-0079294.
2328	ggcacaatgaacggatcacctcgtcaactggtgcaaaagtgtatggttttcaacatcacag	1881	PR	98US-0079656.
1822	ccagcccatgagaggtgagcactatctaaaggtatccaggaatcgtgaagaagctcc	2447	PR	98US-0079664.
2388	ccagcccatgagaggtgagcactatctaaaggtatccaggaatcgtgaagaagctcc	1941	PR	98US-0079689.
1882	tacgagagctctcacagttcaaaaaactgagtcggtgagtcaggtatcccccaagacc	2507	PR	98US-0079728.
2448	tacgagagctctcacagttcaaaaaactgagtcggtgagtcaggtatcccccaagacc	2001	PR	98US-0079786.
1942	ccagaagtaacctaggtcagtgaggggtcttagggaccatggtatagagaggaaacca	2567	PR	98US-0079920.
2508	ccagaagtaacctaggtcagtgaggggtcttagggaccatggtatagagaggaaacca	2060	PR	98US-0079923.
2002	agaaaaagagcagaacaaatcagggtcagaaaaagcaaaagcaaaagcaaaagcaaaaga	2627	PR	98US-0080105.
2568	agaaaaagagcagaacaaatcagggtcagaaaaagcaaaagcaaaagcaaaagcaaaaga	2107	PR	98US-0080107.
2061	agaagaaacagcagaagcagtgctcaggttcaacttgccattcaggt	2674	PR	98US-0080194.
2628	agaagaaacagcagaagcagtgctcaggttcaacttgccattcaggt		PR	98US-0080327.
RESULT 4			PR	98US-0080328.
AAZ33987	AAZ33987 standard; cDNA; 4650 BP.		PR	98US-0080333.
ID	AAZ33987		PR	98US-0080334.
XX	AC		PR	98US-0081049.
XX	AAZ33987;		PR	98US-0081070.
DT	07-DEC-1999 (first entry)		PR	98US-0081071.
DE	Human PRO708 nucleotide sequence.		PR	98US-0081195.
XX	Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;		PR	98US-0081203.
XX	probe; blood coagulation disorder; cancer; cellular adhesion disorder;		PR	98US-0081229.
KW	secreted protein; transmembrane protein; ss.		PR	98US-0081817.
KW			PR	98US-0081838.


```

Db 1909 ctctcagagcagcaaaagacacataattggaagaggaggtacccggctgtaggcttctgtgc 1968
QY 1402 atagccactcttgaaacaaaggaacagtggtgaaggaacacctgtgacacatcactgact 1461
Db 1969 atagccactcttgaaacaaaggaacagtggtgaaggaacacctgtgacacatcactgact 2028
QY 1462 ggtacccactctcattctcactggtcgaagacagattgagaggacattcaactagatg 1521
Db 2029 ggtacccactctcattctcactggtcgaagacagattgagaggacattcaactagatg 2088
QY 1522 gctatgatacttggaagacataaagtgagggtcttcgctcaccgccagtagattttgc 1581
Db 2089 gctatgatacttggaagacataaagtgagggtcttcgctcaccgccagtagattttgc 2148
QY 1582 ataacttgaccccatatcaccaaggaacaaatggctcctggcagcaggtatggga 1641
Db 2149 ataacttgacccc-tatacaccaaggaacaaatggctcctggcagcaggtatggga 2207
QY 1642 tctggaacactgcaatccagtcagccacagagtcagcagcactggaatgtcttacaggaa 1701
Db 2208 tctggaacactgcaatccagtcagccacagagtcagcagcactggaatgtcttacaggaa 2267
QY 1702 atctggctacagcagctgggtcccccctcagctctttcagcaacacctgggacacacggt 1761
Db 2268 atctggctacagcagctgggtcccccctcagctctttcagcaacacctgggacacacggt 2327
QY 1762 ggcaaatgaacggtatcacctctgcaactggaaggaagtgatgcttttcaacatcacag 1821
Db 2328 ggcaaatgaacggtatcacctctgcaactggaaggaagtgatgcttttcaacatcacag 2387
QY 1822 ccgacccatgagaggggtggaacctatctaacaggtatccaggaatcgtgaagaagctcc 1881
Db 2388 ccgacccatgagaggggtggaacctatctaacaggtatccaggaatcgtgaagaagctcc 2447
QY 1882 tacggaggctctcacagttcaacaaactcagtcgaggtcaggtatcccccaagacc 1941
Db 2448 tacggaggctctcacagttcaacaaactcagtcgaggtcaggtatcccccaagacc 2507
QY 1942 ccgaagtaacctagctcaatggagggtctgaggaacctggtatgagaggaaccca 2001
Db 2508 ccgaagtaacctagctcaatggagggtctgaggaacctggtatgagaggaaccca 2567
QY 2002 agaaaaagagccagcaaaatcaggtcagaaaaagcaaaagaaagc-aaaaaaaga 2060
Db 2568 agaaaaagagccagcaaaatcaggtcagaaaaagcaaaagaaagc-aaaaaaaga 2627
QY 2061 agaaagaaagcagaaagcaggtcaggttcaacttgcattcaggt 2107
Db 2628 agaaagaaagcagaaagcaggtcaggttaaacccagcaaaatttggt 2674

```

RESULT 5

AAH46867

ID AAH46867 standard; cDNA; 1800 BP.

XX

AC

XX

DT 25-SEP-2001 (first entry)

XX

DE Human 23553 sulfatase polypeptide coding sequence.

XX

KW Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;

KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.

XX

OS Homo sapiens.

XX

FH. Key

FT 1..1800

FT CDS

FT /*tag= a

FT /product= "23553 sulfatase"

XX

PN W0200155411-A2.

XX 02-AUG-2001.
 XX 31-JAN-2001; 2001WO-US03266.
 XX 31-JAN-2000; 2000US-0495823.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
 XX WPI; 2001-476214/51.
 XX P-PSDB; AAB85483.
 XX Novel human sulfatase polypeptides useful for treating and diagnosing
 XX sulfatase-related disorders such as cerebrovascular diseases, acute
 XX meningitis, multiple sclerosis, degenerative diseases and tumor
 XX Claim 2; Fig 15; 180pp; English.
 XX The invention provides 22438, 23553, 25278 or 26212 human sulfatase
 XX polypeptides and polynucleotides. The sulfatase genes and polypeptides
 XX are useful for treating disorders involving the brain such as
 XX cerebrovascular diseases, infections such as acute meningitis,
 XX demyelinating diseases including multiple sclerosis, degenerative
 XX diseases affecting the cerebral cortex including Alzheimer's disease
 XX and Pick disease, spinocerebellar degenerations including spinocerebellar
 XX ataxias including Friedreich ataxia, and ataxia telangiectasia,
 XX degenerative diseases affecting motor neurons including amyotrophic
 XX lateral sclerosis, inborn errors of metabolism such as leukodystrophies,
 XX toxic and acquired metabolic diseases, including vitamin deficiencies,
 XX and neurocutaneous syndromes (phakomatoses) including neurofibromatosis.
 XX The present sequence represents a human 23553 sulfatase polypeptide
 XX coding sequence.
 XX Sequence 1800 BP; 534 A; 434 C; 423 G; 409 T; 0 other;

Query Match 78.6%; Score 1781.6; DB 22; Length 1800;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1795; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
 QY 324 atggtccacaggggtgtggtggggtccgctccctccctccacagcctgtgtgtgt 383
 Db 1 atggtccacaggggtgtggtggggtccgctccctccctccacagcctgtgtgtgt 60
 QY 384 cctggaagatgctagcaaatggggcggtggtggtggtggtggtggtggtggtggt 443
 Db 61 cctggaagatgctagcaaatggggcggtggtggtggtggtggtggtggtggtggt 120
 QY 444 tatggttacctgtcctggggcagcgttagaagagagagagagagagagagagag 503
 Db 121 tatggttacctgtcctggggcagcgttagaagagagagagagagagagagagag 180
 QY 504 caagctggagaaactagagccagcagcaacttccactccacgccccctcattttc 563
 Db 181 caagctggagaaactagagccagcagcaacttccactccacgccccctcattttc 240
 QY 564 atctagcggatgacagggattagagatggtgtacacagcagcagcagcagcagc 623
 Db 241 atctagcggatgacagggattagagatggtgtacacagcagcagcagcagcagc 300
 QY 624 cctactctgacaagctcgtcgcagagaggttaaaactgagaaactactatgtccagcct 683
 Db 301 cctactctgacaagctcgtcgcagagaggttaaaactgagaaactactatgtccagcct 360
 QY 684 attgcacaccatccaggagtgagtttattactggaagtatcagatacacaccgagctt 743
 Db 361 attgcacaccatccaggagtgagtttattactggaagtatcagatacacaccgagctt 420
 QY 744 caacattctatcatgaagcactaccccaacccaactgtttactctgtgagaatgcaccccta 803
 Db 421 caacattctatcatgaagcactaccccaacccaactgtttactctgtgagaatgcaccccta 480

